Article:
Rodríguez, Raquel Gil, Hedjar, Laysa, Toscani, Matteo et al. (3 more authors) (2024) Color constancy mechanisms in virtual reality environments. Journal of Vision. 6. ISSN 1534-7362

https://doi.org/10.1167/jov.24.5.6

Color constancy mechanisms in virtual reality environments

Raquel Gil Rodríguez* Psychology Department, Justus-Liebig University, Giessen, Germany

Laysa Hedjar* Psychology Department, Justus-Liebig University, Giessen, Germany

Matteo Toscani Psychology Department, Bournemouth University, Poole, UK

Dar'ya Guarnera School of Arts and Creative Technologies, University of York, York, UK

Giuseppe Claudio Guarnera Department of Computer Science, University of York, York, UK

Karl R. Gegenfurtner Psychology Department, Justus-Liebig University, Giessen, Germany

Prior research has demonstrated high levels of color constancy in real-world scenarios featuring single light sources, extensive fields of view, and prolonged adaptation periods. However, exploring the specific cues humans rely on becomes challenging, if not unfeasible, with actual objects and lighting conditions. To circumvent these obstacles, we employed virtual reality technology to craft immersive, realistic settings that can be manipulated in real time. We designed forest and office scenes illuminated by five colors. Participants selected a test object most resembling a previously shown achromatic reference. To study color constancy mechanisms, we modified scenes to neutralize three contributors: local surround (placing a uniform-colored leaf under test objects), maximum flux (keeping the brightest object constant), and spatial mean (maintaining a neutral average light reflectance), employing two methods for the latter: changing object reflectances or introducing new elements. We found that color constancy was high in conditions with all cues present, aligning with past research. However, removing individual cues led to varied impacts on constancy. Local surrounds significantly reduced performance, especially under green illumination, showing strong interaction between greenish light and rose-colored contexts. In contrast, the maximum flux mechanism barely affected performance, challenging assumptions used in white balancing algorithms. The spatial mean experiment showed disparate effects: Adding objects slightly impacted performance, while changing reflectances nearly eliminated constancy, suggesting human color constancy relies more on scene interpretation than pixel-based calculations.

Introduction

Our ability to perceive colors plays a significant role in our daily lives. Color allows us to perceive things quicker and to remember them better (Gegenfurtner & Rieger, 2000; Wichmann, Sharpe, & Gegenfurtner, 2002). Our visual system processes the visible spectrum and interprets the colors of objects based on how they reflect or emit light. Interestingly, we can still identify the same object under different lighting conditions. For example, whether we are in a room with an office lamp or outside during sunset, the pages of a book will appear white. This phenomenon is known as color constancy. It is our ability to perceive the invariant reflectance properties of an object, despite changes in the illuminant color (Helmholtz, 1867; Helmholtz, 1910; Hering, 1920; von Kries, 1923; Judd, 1940). Many of the early studies disagreed on how color constancy was achieved, but there was a consensus that constancy was rather high, at least when the field of view was large and the scene structured (Katz, 1911).

Starting with the experiments of Arend and Reeves (1986), color constancy was investigated in a more quantitative manner, trying to obtain a single number

Citation: Gil Rodríguez, R., Hedjar, L., Toscani, M., Guarnera, D., Guarnera, G. C., & Gegenfurtner, K. R. (2024). Color constancy mechanisms in virtual reality environments. *Journal of Vision*, *24*(5):6, 1–28, https://doi.org/10.1167/jov.24.5.6.

https://doi.org/10.1167/jov.24.5.6 Received September 30, 2023; published May 10, 2024 ISSN 1534-7362

(the color constancy index) that would specify exactly to what degree human observers are color constant: A value of 0 indicates a lack of color constancy, while a value of 1 or 100% indicates perfect color constancy performance. Arend and Reeves pioneered this quantitative approach and initially observed relatively low constancy. Their estimate was as low as 20% under some of their specific settings. However, with the introduction of a variety of different methods over the years, estimates of the constancy index have steadily increased, reaching levels of 80% and beyond (for reviews, see Foster, 2011; Witzel & Gegenfurtner, 2018). The picture that emerges is that constancy can be quite high when the scene is naturally structured and lit by a single illuminant (Kraft & Brainard, 1999; Hansen, Walter, & Gegenfurtner, 2007; Olkkonen, Witzel, Hansen, & Gegenfurtner, 2010; Radonjić, Cottaris, & Brainard, 2015; Radonjić et al., 2018; Hurlbert, Gupta, Gross, & Pastilha, 2019; Gegenfurtner, Weiss, & Bloj, 2024). In many of these experiments, real, physical setups were used, rather than simulations of scenes on computer displays. While this made it possible to show that constancy can be high, it is cumbersome at best and impossible at worst to manipulate the scene experimentally. Recent advances in virtual reality (VR) technology together with three-dimensional (3D) rendering engines have facilitated immersive experiences in different research fields (Cipresso, Wilson, & Soranzo, 2015; Yaremych & Persky, 2019; Shapiro & LoPrete, 2020; Wiesing, Fink, & Weidner, 2020). Here, we use VR to allow for nearly natural conditions, while at the same time maintaining full experimental control. We use a color selection task to study the contribution of three basic mechanisms to color constancy—local contrast, bright is white, and global mean color.

The choice of method is important for these experiments, and numerous methods have been used for measuring color constancy, including achromatic adjustment (Kraft & Brainard, 1999; Hurlbert et al., 2019), color selection (Radonjić et al., 2015; Gegenfurtner et al., 2024), color categories (Olkkonen et al., 2010), and illumination discrimination (Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert, 2014; Radonjić et al., 2018). In achromatic adjustment, the color of the target object is changed until it appears gray to the observer or is adjusted until it matches a reference one. In a color selection task, participants choose the color-constant object from a group of competitors. In a color categories task, a color patch is selected from a given chart system, such as from Munsell color chips. Unlike the previous tasks, illumination discrimination involves identifying the color of the light source instead of the objects and is an indirect measure of color constancy. Observers focus on the illuminant to make this determination. For our experiment, we opted for a color selection task because in a virtual reality environment, it encourages participants to look around the scene for objects that are already integrated into the environment.

In terms of how color constancy is achieved, researchers have identified and studied three primary mechanisms: (a) local surround, which states that color changes across local edges (for example, between object and background) are relatively stable across illuminant changes (Wallach, 1948; Valberg & Lange-Malecki, 1990; Foster & Nascimento, 1994; Nascimento & Foster, 2000); (b) maximum flux, or "bright is white," in which we adapt to the brightest area of a scene (Land & McCann, 1971; Land, 1986; Cataliotti & Gilchrist, 1995; Gilchrist et al., 1999); and (c) spatial mean, or "gray world," which refers to adaptation to the average color of a scene (Buchsbaum, 1980; Brainard & Wandell, 1986). Here, we examine these three classical color constancy mechanisms in an immersive environment considering five illuminations: three on the daylight locus (neutral, blue, yellow) and two in the orthogonal direction (red, green). To silence the local surround mechanism, we place the target object on a leaf that has a constant pink-rose chromaticity under all illumination conditions. To silence the maximum flux mechanism, we place a bright object in the scene that has constant neutral chromaticity under all illuminations. It is more difficult to keep the spatial mean chromaticity constant under illumination changes. To silence this mechanism, we define two different approaches: (a) adding new objects into the scenes and (b) modifying the reflectances of the current objects.

These three mechanisms were investigated earlier by Kraft and Brainard (1999) in a setup with real objects, and they measured the contribution of various combinations of the cues to color constancy. Their study was seminal in showing that constancy crucially depends on the quality of visual information that is available. However, it was not possible for them to separately and independently silence the three cues under arbitrary illumination changes. Here, we create two scenes, one indoor and one outdoor, and conduct a color selection experiment for the three mechanisms together with a baseline experiment, where all cues are present. Observers choose among five competitor objects to match the color of the target object (achromatic), shown at the beginning of the experiment under neutral illumination. The scenes are modeled in Autodesk 3ds Max, rendered using the real-time gaming engine of Unreal Engine (v4.27.1), and finally visualized in the head-mounted display HTC Vive Pro Eye.

Our study reveals that color constancy performance was high in the "all cues present" baseline condition, consistent with previous studies on color constancy in complex real-world scenes (Kraft & Brainard, 1999; Gegenfurtner et al., 2024). However, selectively silencing each classic cue had differing effects on constancy.

On the one hand, performance significantly dropped for all colored illuminations when local surround was eliminated as a cue. This was particularly so under the green illuminant, indicating significant interactions between the greenish illumination and the rose-colored local context, two colors on opponent poles of color space. On the other hand, the maximum flux mechanism had minimal impact on performance, and the performance drop was not statistically significant. This was surprising because many computational algorithms for white balancing use the brightest (or white) patch in the scene (Land & McCann, 1971; Land, 1986). The most intriguing finding was from the spatial mean experiment, where the two approaches we took had vastly different results. While adding new objects to balance the average color in the scene had a small impact on performance, changing reflectances of existing objects almost completely abolished constancy. This indicates that human color constancy is not achieved using pixel-based computations but relies on scene segmentation and interpretation. Our results encourage an elaboration of the classic color constancy cues for real-world scenes.

Figure 1. The two 3D rendered environments in Unreal Engine: indoor and outdoor, both under neutral illumination.

Method

In this section, we present the setup and the experiments we conducted to measure color constancy performance.

Stimuli

VR environment

We modeled two scenes using Autodesk 3ds Max with the goal of creating two realistic and natural settings: an indoor office-home environment and an outdoor forest scene rendered using the gaming engine Unreal Engine (v4.27.1). Figure 1 shows the two scenes from one angle. The indoor scene contains a desk, an office chair, a lounge chair, a door, a bookshelf, paintings, and curtains. There are several small objects placed around the scene, including plants and books, which have a variety of material properties and reflectances, including specularities. We used a single point light source placed just above the observers' VR actor in gameplay (center of the room above office chair), such that even if they looked up, they would not see the light source.

The outdoor scene spans a larger area in VR space, containing trees, a large cliff, a lake, rocks, moss, grasses, and flowers. There were minor ripples across the lake water; otherwise, all other objects were static. The light sources were a directional light, pointed toward the large cliff, along with a skylight set to the same color as the directional light. We additionally changed the color of the sky by multiplying its default material colors (bluish with white clouds) by the illuminant color, such that the sky color (that is, its reflected light) was influenced by the color of the illuminant. The water in the lake was translucent and its color was shifted toward that of the illuminant. It reflected the sky as well as other surroundings (e.g., trees) at its surface.

To render the scenes, we make use of the two-step Photon Mapping (Jensen, 2001) algorithm available in Unreal Engine (*Lightmass*). The first step, computed offline, performs a lighting simulation by tracing packets of photons emitted by a light source, with each photon carrying a fraction of the power of the light. As photons scatter within the scene and hit nonspecular surfaces and are either scattered or partially absorbed, their energy, locations, and directions are stored in photon maps (*surface lightmaps*), thus enabling subsequent computation of reflected radiance at any point in the scene. As dense lighting samples are collected throughout the volume of the scene (*Volumetric Lightmaps*), lighting data can be interpolated and used to light dynamic objects. Mirror reflections are computed in the second step, which make use of Monte Carlo ray tracing to render the final images. Since most of the illumination information is computed offline during the first step, the computational cost of the rendering step is significantly lower than pure Monte

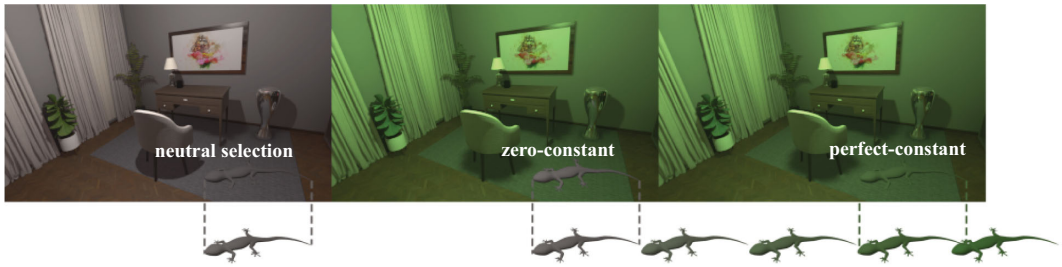

Figure 2. Indoor scene under neutral (left) and green illuminants (middle and right). The lizard on the left corresponds to the perfect/zero-constant lizard under the neutral illuminant. The five lizards on the right correspond to the five color competitors under the green illuminant: zero-constant match, two samples, perfect-constant match, and over-constant match.

Carlo ray-tracing approaches, albeit at the cost of higher storage requirements. Overall, photon mapping tends to produce low-frequency noise, rather than the high-frequency artifacts introduced by variance in Monte Carlo ray tracing. A potential downside of photon mapping is the bias introduced in rendered images, due to the kernel used in photon density estimation from stored data (Schregle, 2003). However, since the technique is consistent (Jensen, 2001), we reduced the bias by increasing the number of photons traced offline during the first step.

Illuminants

We rendered each scene under each of five illuminants. We used the same illuminants as Aston, Radonjić, Brainard, and Hurlbert (2019), with three along the daylight axis (neutral, blue, yellow) and two orthogonal (red, green). The CIE *xy* chromaticites are listed here: neutral (.31, .33), blue (.25, .26), yellow (.39, .39), red (.32, .26), and green (.30, .38). When working with the engine, we need to specify the *RGB* values for the light and its intensity. For conversion to *RGB*, we set the luminance (Y value) equal to 30 cd/m^2 and used the precomputed calibration from Gil Rodríguez et al. (2022). Additionally, in Unreal Engine, one can specify the intensity of the light sources. In the indoor scene, we set the intensity to 16.58 cd. In the outdoor scene, the directional light source was set to 15 lux.

Color competitors: Lizards

We used a lizard-shaped mesh for our test stimuli (see Figure 2). In the beginning of each VR session, observers were presented with the reference stimulus, an achromatic lizard, which was placed in the scene under the neutral illuminant. The reflectance of the lizard was set to $RGB = [0.146, 0.145, 0.147]$, which resulted in a reflected light of chromaticity values $(x, y) = (0.31, 0.33)$ under the neutral illuminant. In the indoor scene, this lizard sat on the brown desk in front of the observer, and in the outdoor scene, it sat on the moss in front of the cliff (see Figure 4 under "Lonely Lizard").

During the experimental trials, five lizards were placed in the scene, each one a different color competitor. Color competitors were chosen based on the illuminant in the scene (red, green, blue, yellow, or neutral). *RGB* reflectances and reflected light were calculated using the following formula:

$$\text{Reflected}_{RGB} = \text{Reflectance}_{RGB} \times \text{Illuminant}_{RGB}. \quad (1)$$

"Reflectance" is defined as the proportion of light of each channel that an object reflects and is a property of an object regardless of illumination. "Reflected light" is defined as the light that is reflected from an object and is a combination of the reflectance and the illumination. We computed the five competitors for the colored illuminations as follows. Figure 2 shows the color competitors under the green illuminant:

(1) The *perfect-constant match* (or *reflectance match*) had the same reflectance *RGB* as the reference lizard.
(2) The *zero-constant match* (or *tristimulus match*) reflected the same light under all illuminants as the reference lizard under the neutral illuminant.
(3) The reflectances of two other competitors were evenly spaced between the zero-constant and perfect-constant matches in CIELAB color space.
(4) The *over-constant match* was placed beyond the perfect-constant match at a distance 25% of that between the perfect- and zero-constant matches.

For the neutral illuminant, the reference lizard color served as both the perfect-constant and zero-constant match. We added two competitors in the blue direction, one with the same reflectance as the zero-constant match in the yellow illuminant and another halfway between in CIELAB, and two competitors in the yellow direction, computed in the same way.

Due to the large range of depths in the outdoor scene, lizard mesh sizes (in VR world units) were adjusted depending on their location in the scene, such that their degrees of visual angle varied between ~8° and 21°. The indoor scene had very little depth

variation, so lizards in all locations were the same size in VR world units (∼18°).

Color calibration

Color vision research in VR requires precise calibration of display devices (Barnard, Cardei, & Funt, 2002). Several studies have focused on calibrating different headsets using various software, such as the HTC Vive Pro Eye and Pimax 5k+ with Unreal Engine (Clausen, Fischer, Furhmann, & Marroquim, 2019; Toscani et al., 2019; Zaman, Sarker, & Tavakkoli, 2023). Recent work also compares calibrations using different rendering software and shaders (Murray, Patel, & Wiedenmann, 2022; Díaz–Barrancas, Gil Rodríguez, Aizenman, Bayer, & Gegenfurtner, 2023). VR techniques have been shown to be suitable in principle for studying color constancy (Díaz-Barrancas, Cwierz, Gil Rodríguez, & Pardo, 2022; Gil Rodríguez et al., 2022). Here we use VR to study specific cues for color constancy in indoor and outdoor environments.

The colors of the illuminants and lizards were calibrated according to the procedure described by Gil Rodríguez et al. (2022). The head-mounted display (HMD) used was an HTC Vive Pro Eye, which has two AMOLED displays each with a resolution of 1,440 × 1,600 pixels with 24-bit resolution, a maximum field of view of 110°, and a refresh rate of 90 Hz.

Briefly, we measured the primaries of the HMD using a Konica Minolta CS-2000A spectroradiometer. Using Unreal Engine, we rendered a simple scene: a room with black-and-white matte checkerboard walls (R, G, $B = 0$ and R, G, $B = 1$), a point light source (R, G, $B = 1$), and a matte uniform surface placed under the light. The center of the HMD pointed to the surface, and we measured the light emitting from the center of the HMD while varying the R, G, and B channels of the surface's reflectance. All postprocessing and tonemapping used by the engine were disabled, which left us with a linear relationship between channel bit value and luminance, with luminances saturating after a certain value (Toscani et al., 2019; Gil Rodríguez et al., 2022). The dynamic range measured under this calibration setup was 1:20,000. This calibration was validated by setting the surface's reflectance RGB to 100 different colors spanning the RGB gamut and measuring the emitted light (please refer to Díaz-Barrancas et al., 2023 for more details). Figure 3 shows the difference in CIEDE2000 (ΔE00) units between the nominal and measured values of these 100 colors; note that the ΔE00 never exceeds 2, which is below the minimal just-noticeable-difference threshold for ΔE00 units (Sharma, Wu, & Dalal, 2005). During gameplay of our experiment, all postprocessing was disabled to ensure accurate color calibration (Gil Rodríguez et al., 2022).

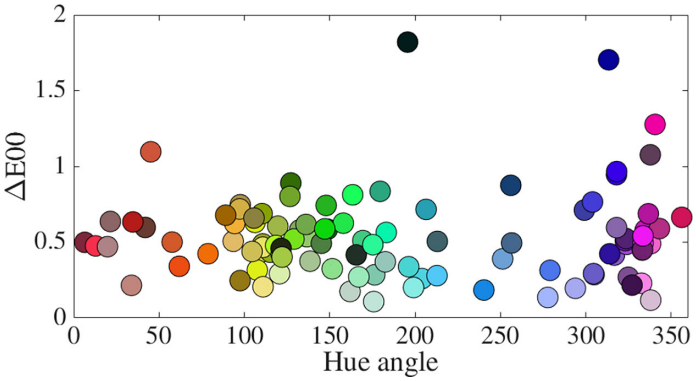

Figure 3. Color validation of the headset calibration: ΔE00 values between nominal colors and measured colors. See text for details.

Observers

Ten naive observers completed the experiment in the indoor scene, and 10 separate naive observers completed the experiment in the outdoor scene. All observers gave informed consent and had normal color vision as assessed by the Ishihara Color Vision Test. All observers were recruited from the university and understood and spoke English fluently.

Procedure

Observers wore the HMD and were holding one HTC Vive controller while seated in a rolling chair for the duration of the experiment. At the beginning of each VR session, once they were comfortable in the headset, they completed a 5-point eye-tracking calibration built into the SteamVR platform used to control the headset (ViveDevelopers). For the calibration, we first adjust the headset position and the interpupillary distance, and then observers follow a dot on the screen, which moves to five different locations, with their eyes. We receive feedback from SteamVR on whether the calibration is successful. If it is, the observer proceeds with the experiment; otherwise, they repeat the calibration (Imaoka, Flury, & de Bruin, 2020). Analysis of this eye movement data is not presented in this publication.

Figure 4 depicts the timeline of the experiment within the VR environment. Observers first sat in a room with multicolored walls on which instructions in English were written. Observers stayed in this room for a minimum of 30 seconds before the words "PRESS the TRIGGER" appeared, allowing them to move on to the experiment. Once they pressed the trigger on the VR controller, their VR actor was transported to one of the scenes (indoor or outdoor, depending on the observer group) cast under the neutral illuminant. The achromatic reference lizard lay in the scene and

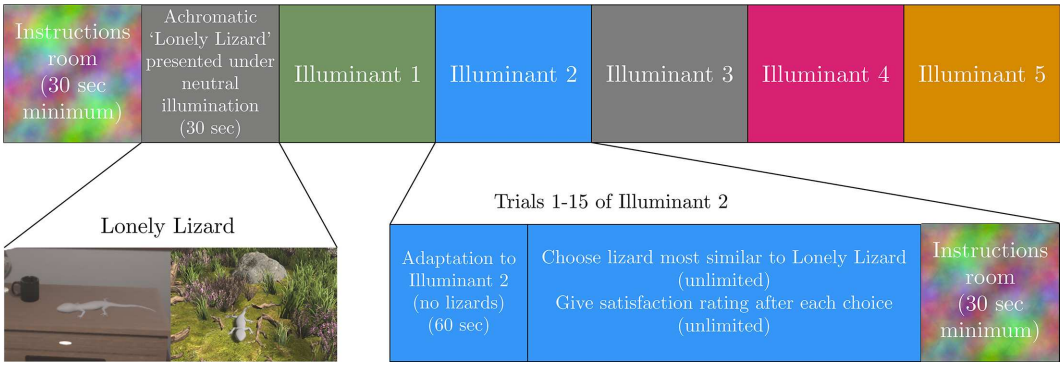

Figure 4. Timeline of a single VR session. Observers first sat in the Instructions room for at least 30 seconds. After pressing the trigger, they moved on to the presentation of the "Lonely Lizard," an achromatic lizard sitting in the scene under the neutral illuminant. Then they were transported to the scene presented under one of the five illuminants to begin the adaptation period. During this time, no lizards were present in the scene. The adaptation period lasted 60 seconds and then the trials began. Five lizards were scattered around the scene, and observers chose the one that appeared most in color like the "Lonely Lizard." After each lizard selection, the observer gave a satisfaction rating. Then the next trial began. At the end of the 15-trial illuminant block, the observer was transported back to the Instructions room to begin the next illuminant block. Illuminant block order was randomized for each session.

Figure 5. Ten lizard locations in the indoor scene. Five are located in front of the chair and five more behind. The colors of the lizards are only for illustration purposes.

the text "Lonely Lizard" appeared floating for a few seconds in the scene before disappearing. After 30 seconds, the Lonely Lizard disappeared. The illuminant abruptly switched to one of the five illuminant colors (blue, green, neutral, yellow, or red), beginning the adaptation period for that illuminant. The text "Look around" appeared in white for a few seconds. Observers were encouraged to rotate in their chair and view the entire scene. After 60 seconds of adaptation, the trials began. Five lizard competitors, each with a different reflectance, appeared around the scene. Observers were told to find all five lizards and then select the one that appeared most like the "Lonely Lizard" they saw at the beginning of the session. They were free to rotate and move in their chair. In order to explore the scene, they had to rotate to change their viewpoint; however, no translation was required to see the relevant areas of the scene. They made their selection by pointing the VR controller, which constantly emitted a turquoise "laser," at the chosen lizard and pressing the trigger. Afterward, a widget appeared on the screen and they were asked to indicate how satisfied they were with their choice on a Likert scale, again using the controller: strongly dissatisfied, dissatisfied, neutral, satisfied, strongly satisfied. The next trial began immediately afterward. Under a single illuminant, observers completed 15 trials before moving on to the next illuminant block. Within a session, trials were blocked by illuminant, with the order randomized, and all five illuminants were tested. In between illuminant blocks, the observers' VR actor moved back to the instructions room for 30 seconds. One VR session lasted about 45 minutes.

On every trial, the five lizards appeared in five locations randomly selected from a total of 10 (indoor) or 9 (outdoor) possible locations in the scene. In the indoor scene, these locations spanned the entire room except the ceiling (Figure 5). In the outdoor scene, the locations spanned most of the scene but excluded distant locations (Figure 6). All lizards were perched on an object (i.e., none were hanging in the air). Additionally, at the beginning of every trial, the view of the VR actor was oriented to a random direction,

Figure 6. Nine lizard locations in the outdoor scene. Six are located on the land, cliff, and tree, and three are around and on top of the water. The colors of the lizards are only for illustration purposes.

in order to oblige the observers to view different parts of the scene. In the indoor scene, the VR actor always hovered over the office chair, but the rotation of the viewpoint could change to one of three angles on every trial. In the outdoor scene, we used three possible VR actor positions and viewpoint rotations.

Cue silencing

Our aim in this experiment was to examine the role three classic constancy cues play on color constancy indices. To do this, we selectively silenced a different constancy cue while keeping all other cues present. As a baseline, we also measured color constancy indices when all cues were present. Sessions were blocked by cue (Baseline [all cues present], Local Surround, Maximum Flux, Spatial Mean: Adding Objects, and Spatial Mean: Changing Reflectances), and observers completed two sessions of each cue, for a total of 30 repetitions per condition. Cue silencing did not affect presentation of the reference lizard, which was always done under a neutral illuminant with all cues present. Sessions were randomized with one caveat: The two Spatial Mean cues were added to the experiment later and thus were completed last. Below we explain the changes we made to the scenes in order to silence each cue. All other aspects remained the same.

Local surround

To silence the local surround cue, we chose to place the lizards on a natural object slightly larger than the size of the lizard. We used a taro leaf in the indoor scene and a castor leaf in the outdoor scene. We set the leaf material's shading model to "Unlit" and defined the emissive color of the leaf as $RGB = [0.425, 0.183, 0.187]$, a pinkish rose-like color. This means that the leaf color is self-illuminating (but is not a light source) and is not affected by the illuminant, so across all illuminants, it reflects the same light. To choose the leaf color, we wanted to avoid a color close to the illuminants or the competitors. The rose color lies between the reddish and yellowish illuminants in CIELAB space at equal chroma (where $R, G, B = 1$ is defined as $a^*, b^* = 0, 0$). Figure 7 illustrates this manipulation in the outdoor scene under each illumination.

Maximum flux

The maximum flux cue presupposes that the brightest reflected light in the scene is recomputed by the observer as white. To silence this cue, we set an object in the scene to always reflect the brightest light and to be unaffected by the illuminant. In the outdoor scene, we placed a matte uniform square sheet (23°) on the cliff that conformed to the cliff's shape (modeled in Autodesk 3ds Max). The *RGB* reflectance of the sheet under each illuminant was chosen such that the *RGB* reflected light was 1 (see Equation 1). Figure 8 shows an illustration demonstrating how this mechanism was silenced. In the indoor scene, we used two lampshades (9–10°) as the maximum flux cue (on opposite sides of the room). Since the lampshades had a cloth fabric texture, we set its reflected light color in a similar way as for the leaves for the local surround cue: We defined the material as "Unlit" and set the default base color instead as the emissive color.

Spatial mean: Adding objects or changing reflectances

The spatial mean cue presupposes that an observer will take the spatial average color of the scene and normalize it to gray. In order to silence this cue, we needed to equate the average color of the scenes across illuminants; this is tricky because an illuminant globally shifts the average color of the scene toward the color of the illuminant. To silence this cue, we needed to shift the spatial mean under each chromatic illuminant to that under the neutral illuminant. We took two approaches:

(1) Adding new objects to the scene
(2) Changing the reflectances of objects present in the scene

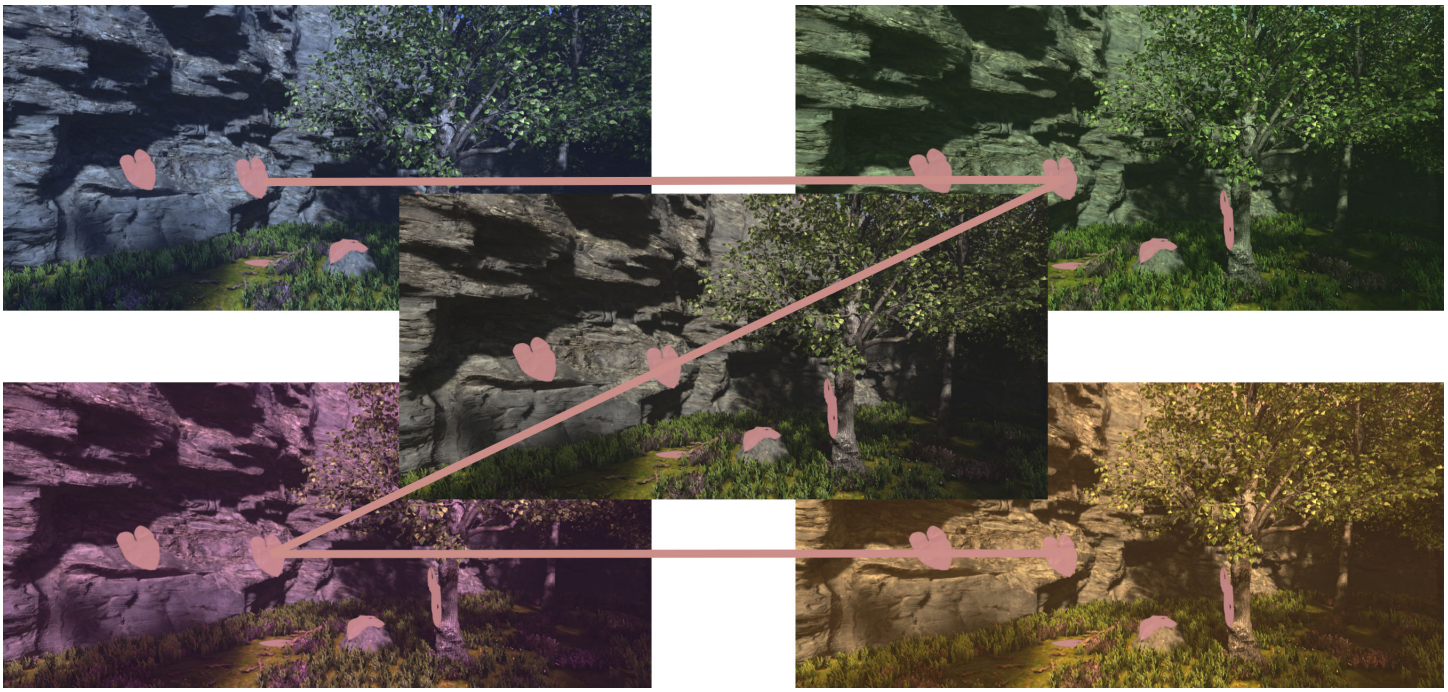

Figure 7. Illustration of the local surround cue manipulation in the outdoor scene. The scene under the neutral illuminant is in the center. Clockwise starting from the top left are the following colored illuminants: blue, green, yellow, red. Each lizard (not shown) is placed on a rose-colored leaf reflecting the same light under all illuminants (see text for details). A thick line matching the color of the leaves connects one leaf across all illuminant images to show that they reflect the same light.

In the first approach, we added new objects ("moon-like" spheres in the outdoor scene and picture frames and foliage in the indoor scene; see Figure 9) whose reflectances lay opposite the illuminant color away from the neutral point in *xy* chromaticity space. For example, under the blue illuminant, the new object reflectances were yellowish and, under the yellow illuminant, bluish. In the indoor scene, the frames had a default texture and material. The frame itself was semi-glossy (with a specularity of 0.5) and the picture inside of the frame was matte and uniform. Both frame and picture were shifted in color toward the opposite of the illuminant color. For the foliage (ivy plant), we shifted the tint of the albedo also toward the opposite of the illuminant color. In the outdoor scene, the spheres had a concrete, "moon-like," gray default material, so we also shifted the tint of its albedo toward the opposite of the illuminant color.

In the second approach, we shifted the reflectances of almost all objects toward the color opposite the illuminant in the same manner as the previous approach (see Figure 10). We applied this shift to almost all colors in the scenes, except for objects smaller than the office chair in the indoor scene and for the flowers and small grasses in the outdoor scene.

We measured the average scene color by taking images of each scene under one illuminant and averaging. We used a Radiant Vision Systems I29 colorimeter calibrated to the HMD's primaries (ProMetric Software, Radiant Vision Systems, 2023) and fitted with an AR/VR lens specifically designed for use with HMDs. We rotated the camera viewpoint in the VR scene around the z-axis to eight angles (45° apart) for the indoor scene and four angles (45° apart) for the outdoor scene and took pictures. The colorimeter has a resolution of 6,576 × 4,384 and gives us CIExyY values at each pixel; we averaged together the XYZ values across all the images. We provide a sample image acquired by the colorimeter I29 (indoor baseline under yellow illuminant, luminance only shown) on the left side of Figure A.1 in the Appendix. For both approaches, we continuously measured and remeasured the spatial mean color using the colorimeter while adjusting the colors until the values under the colored illuminants roughly equated the spatial mean under the neutral. Figure A.1 shows this by plotting the spatial mean color of the indoor (upper) and outdoor (lower) scenes in CIExy chromaticities under each of the illuminants with "all cues present" (Baseline) as color-coded squares. The circles and triangles correspond to Spatial Mean: Adding Objects and Spatial Mean: Changing Reflectances, respectively, for both indoor and outdoor scenes.

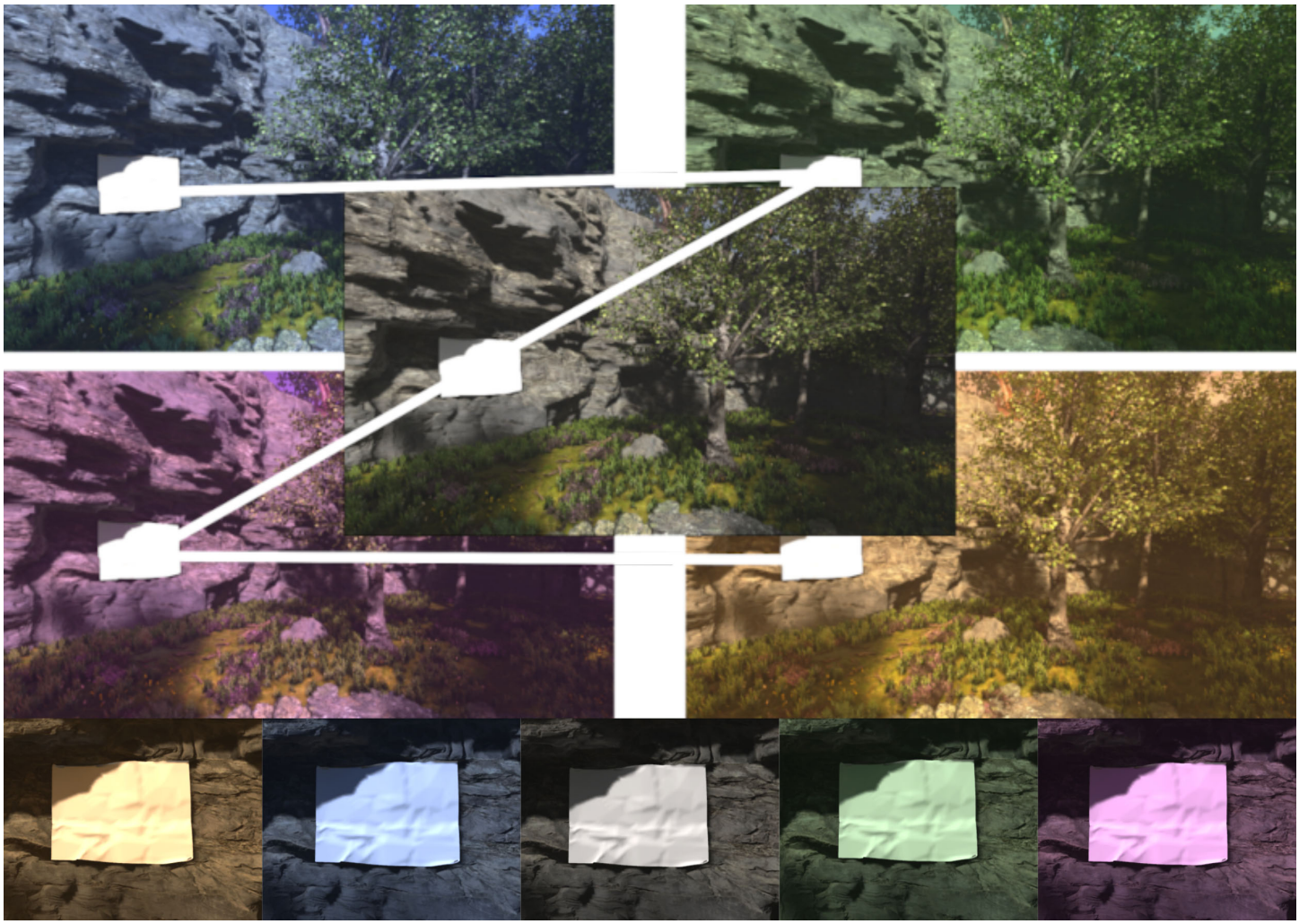

Figure 8. Illustration of the maximum flux cue manipulation in the outdoor scene. Layout of the upper part of the figure is the same as Figure 7. A thick line matching the color of the white sheet connects the cue across illuminant images to demonstrate that they reflect the same light. In the lower part of the figure, we contrast the top images by depicting how the sheet would look if we did not control its reflected light in such a way.

Analysis

One observer from the indoor scene and two from the outdoor scene did not complete the Spatial Mean: Changing Reflectances condition, and two observers from the indoor scene and one from the outdoor did not complete the Spatial Mean: Adding Objects condition. Before any analysis, we removed all trials where observers marked their satisfaction as "strongly dissatisfied" or "dissatisfied." We discarded 16.5% of trials from the indoor scene and 4.1% of trials from the outdoor scene due to low satisfaction ratings.

Since participants could only choose discrete competitors for their lizard match, we chose to use an adaptation of maximum likelihood difference scaling (MLDS) (Maloney & Yang, 2003; Knoblauch & Maloney, 2012), implemented by Radonjić et al. (2015) and Radonjić, Cottaris, and Brainard (2015), to determine their exact match. This method supposes that the five competitors are not laid out in each observer's perceptual space as they are in CIELAB color space and finds an arrangement that maximizes the likelihood of an observer's choice matrix. Briefly, we created a matrix of paired comparisons for each condition consisting of all possible pairs of the five competitors (10 unique pairs), where in each cell, we computed the proportion of times the participant chose one in the pair over the other.

Figure 11 shows an example paired comparison matrix of Observer 4 in the Maximum Flux condition under the red illuminant in the indoor scene. For the analysis, we used the same restrictions as Radonjić et al. (2015). The position of the zero-constant match was set to 0, and the standard deviation of the Gaussian noise applied to the perceptual position of every competitor was set to 0.1. The order of the competitors

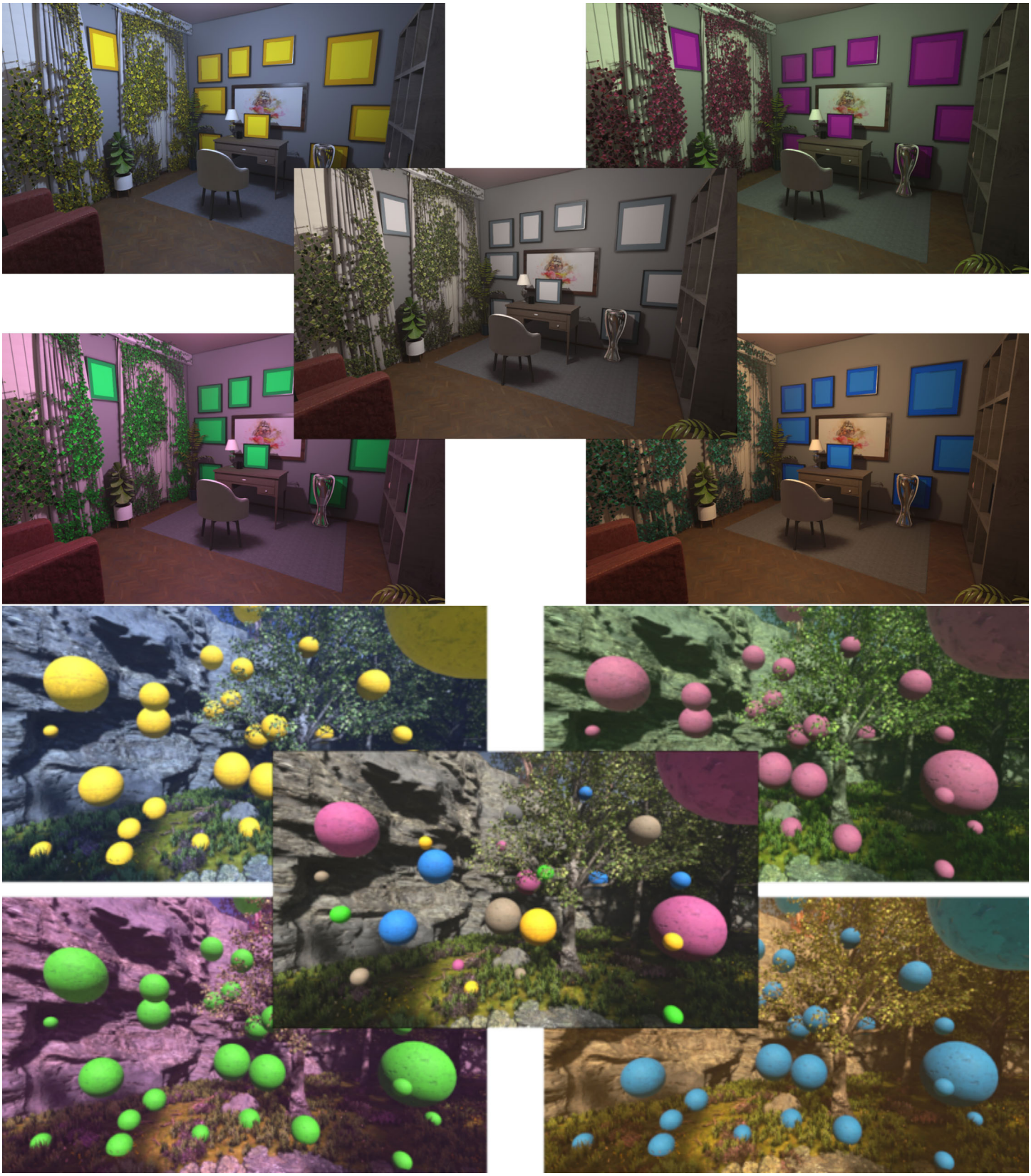

Figure 9. Illustration of the spatial mean cue manipulation by adding new objects to the scene, for the indoor (top) and outdoor (bottom) scenes. Layout of each scene in the figure is the same as in Figure 7. As described in the text, new objects were added to the scene, which balanced the spatial mean color to be neutral.

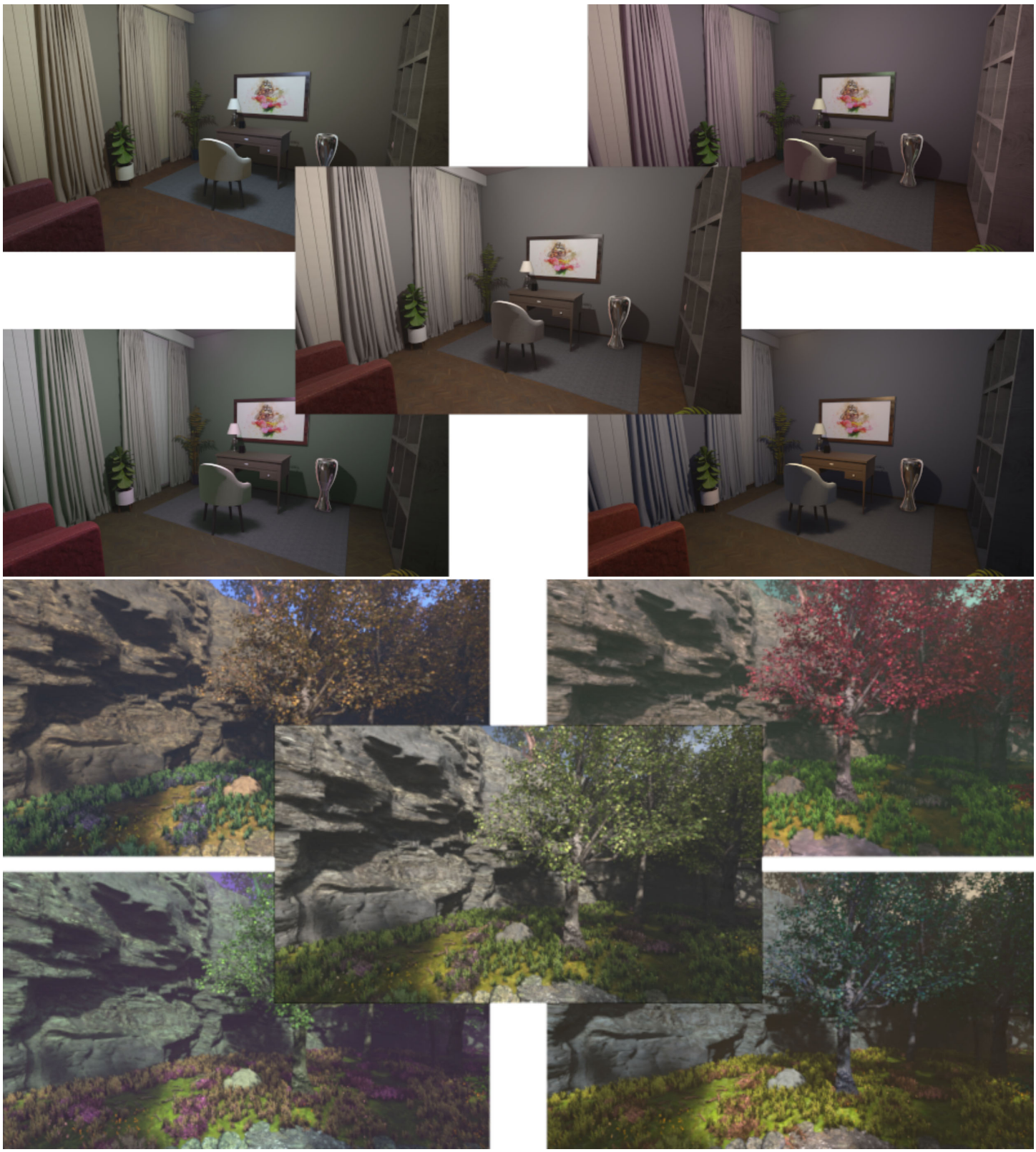

Figure 10. Illustration of the spatial mean cue manipulation by changing the reflectances of objects already present in the scene for the indoor and outdoor scenes. Layout of the figure is the same as Figure 9. As described in the text, we changed the reflectances of almost all objects in the scene in order to balance the spatial mean color to be neutral.

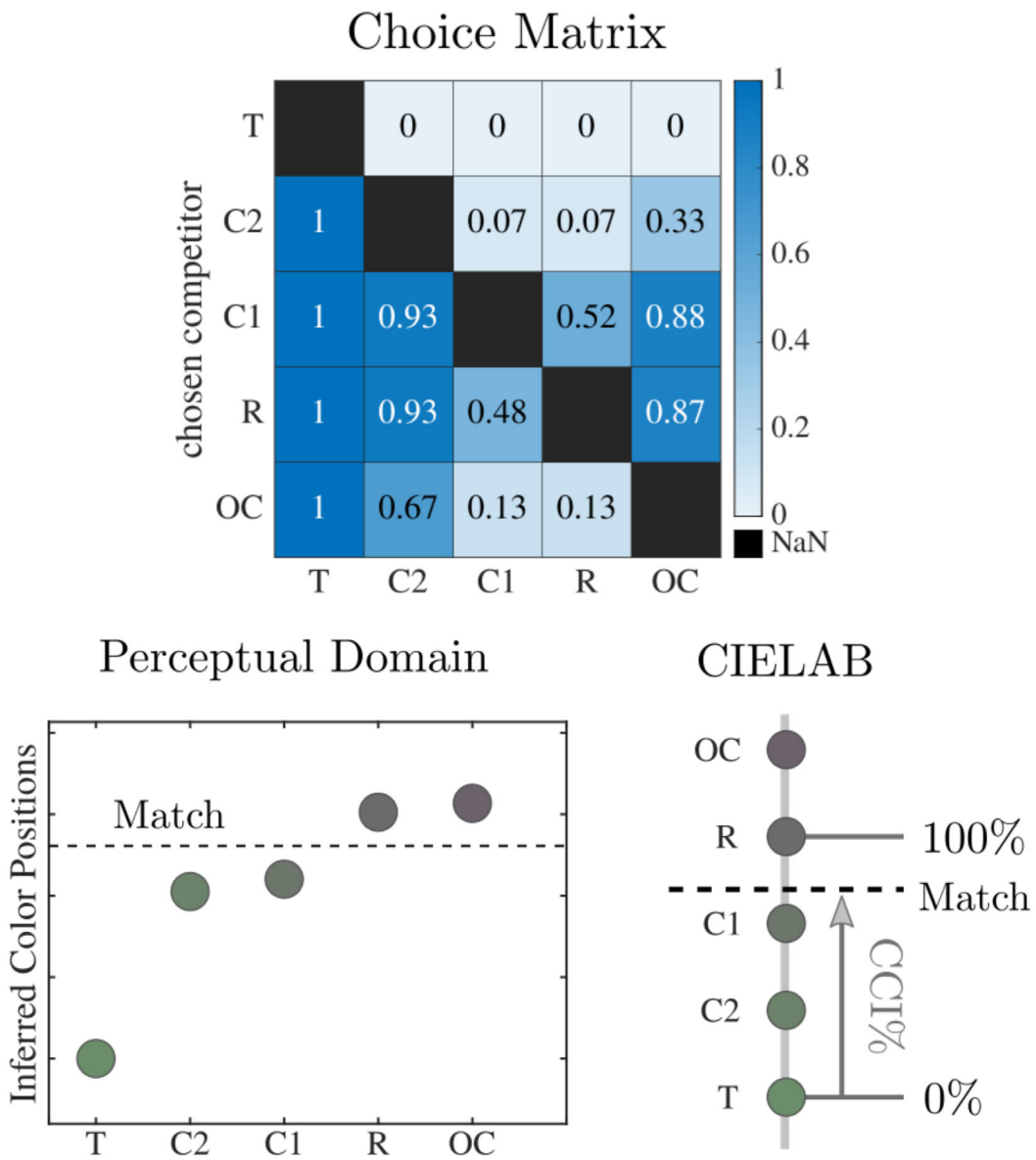

Figure 11. Choice matrix (top) of one sample observer (#4 during the Maximum Flux cue condition under the red illuminant in the indoor scene). Rows represent the chosen competitor—T: tristimulus or zero-constant, R: reflectance or perfect-constant, C1,C2: two competitors between T and R, and OC: over-constant. Each cell notes the proportion of times that competitor was chosen over one other competitor and is color-coded from white (none of the time) to blue (every time). The matrix is symmetrical. Once we find the best competitor and match positions using the modified MLDS algorithm (bottom left), we recover the position of the observer's match in CIELAB space via interpolation (bottom right). See text for more details. Note that competitor colors are defined in the reflectance domain.

was preserved with a minimum distance of 0.025 between competitors. We then iteratively searched for the best-fitting arrangement.

Across all observers and conditions, the R^2 value between the observed proportions and predicted probabilities following the MLDS analysis had an average of 0.99 (95.7% of all R^2s were above 0.95, after discarding two very poor fits with R^2 values below 0.1, both from the outdoor scene), indicating that the algorithm well predicted observers' choices.

Once the location of the observer's match is found in their perceptual space (Figure 11, bottom left), the match can be transformed into CIELAB by preserving the same proportions as in the perceptual space (Figure 11, bottom right). From now on, please note that we defined all the competitors and the estimated matches in the same space by computing their "reflected light" in CIELAB using the white of the monitor as the white point. While the white point of an opponent color space such as CIELAB is typically corrected to

be the color of the illuminant, here we wanted to use the same physical space across all illuminants such that the position of the perfect-constant match would be constant in order to remove any assumption that the observers fully adapted to each illuminant.

Computing the color constancy index (CCI) for the colored illuminants is relatively straightforward: Matches at the zero-constant match had a CCI of 0%, matches at the perfect-constant match had a CCI of 100%, and matches at the over-constant match had a CCI of 125%. For all other matches, CCIs were scaled accordingly in CIELAB. However, it is possible that there are differences between each observer's memory of the color of the reference lizard throughout the session and the actual reference lizard color, presented to them at the beginning of the session. So for each observer, we recalculated the position in CIELAB of a 100% and 0% CCI based on their average match across all conditions under the neutral illuminant. The 0% CCI point is defined as the reflected light from the match under the neutral illuminant. Then, from Equation 1, we can compute its reflectance as

$$\text{Reflectance}_{RGB} = \text{Reflected}_{RGB}/\text{Illuminant}_{RGB}. \quad (2)$$

The 100% CCI point is calculated from this "Reflectance" value using Equation 2 and applying again Equation 1, so we can get the "Reflected" value under each illuminant. Observer CCIs were then calculated from these new points.

For all statistical analyses, we fitted linear mixed-effects models (LMMs) using the *nlme* package (Pinheiro, Bates, & R Core Team, 2023) in the R programming environment (R Core Team, 2023). All LMMs were fitted using the restricted maximum likelihood method. We ran analyses of variance (ANOVAs) using the built-in R package stats. All post hoc contrasts were performed using the *emmeans* package (Lenth, 2023).

Results

Baseline condition: All cues present

Figure 12 plots the color constancy indices of each participant for the baseline condition, in which all cues are present. In line with our earlier experiment (Gil Rodríguez et al., 2022), the degree of constancy under these conditions was very high. We fitted a linear mixed-effects model with *illuminant* as a fixed-effects factor and *observer* as a random-effects factor and performed a one-way ANOVA. We found no difference between CCIs of illuminants in the indoor scene, with a mean of 96.7% and a standard deviation of 18.9%

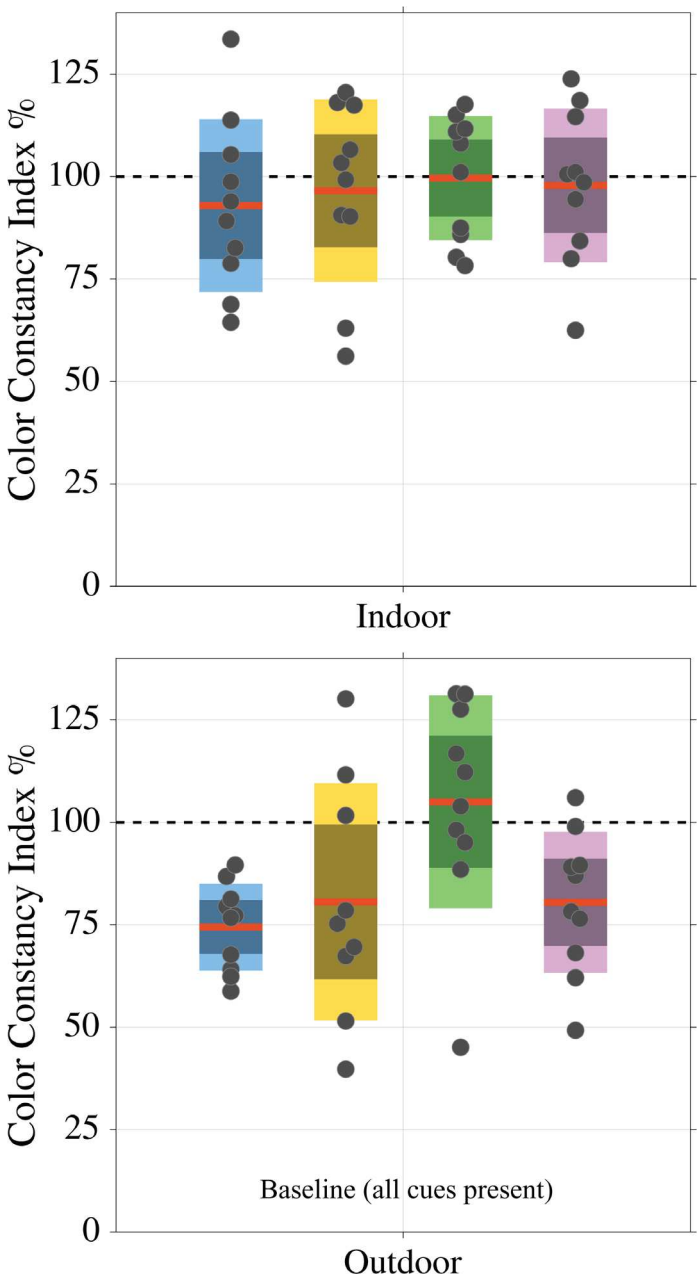

Figure 12. Color constancy indices under the colored illuminants in the baseline condition (all cues present) for both scenes. Each dot represents a participant. Note that there were different sets of participants for each scene. The mean across participants for a given illuminant is marked with a red line. The illuminants are color-coded. The darker region denotes ±1.96 standard errors of the mean and the lighter region denotes ±1 standard deviation from the mean. The dotted line represents perfect color constancy (100%).

across illuminants ($F(3, 27) = 0.38$, $p = 0.77$). However, in the outdoor scene, CCIs for the green illuminant, 105.0% on average, were significantly different from the blue and the red, which had average CCIs of 74.4% and 80.4% ($F(3, 26) = 5.07$, $p = 0.007$; with Tukey-honestly significant difference (HSD) corrections

	Indoor			Outdoor		
Silenced cue	**Est. diff from baseline**	**t-stat**	**p-val**	**Est. diff from baseline**	**t-stat**	**p-val**
Local Surround	−22.62	−4.89	0.00	−14.86	−4.14	0.00
Maximum Flux	−6.79	−1.47	0.14	−0.86	−0.24	0.81
SM: Adding Objects	−20.30	−4.10	0.00	−9.65	−2.61	0.01
SM: Changing Reflectances	−66.82	−14.00	0.00	−66.06	−16.88	0.00
Contrasts within cue (illums)	**Estimate**	**t-stat**	**p-val**	**Estimate**	**t-stat**	**p-val**
Local Surround : blue	4.28	0.78	0.90	6.86	1.25	0.62
Local Surround : yellow	1.70	0.31	1.00	7.37	1.30	0.59
Local Surround : green	−10.66	−1.94	0.20	−23.74	−4.32	0.00
Local Surround : red	4.67	0.85	0.87	9.51	1.67	0.34
Maximum Flux : blue	1.39	0.25	1.00	−7.24	−1.33	0.56
Maximum Flux : yellow	0.73	0.13	1.00	1.06	0.19	1.00
Maximum Flux : green	2.32	0.42	0.99	3.51	0.64	0.95
Maximum Flux : red	−4.44	−0.81	0.89	2.67	0.49	0.98
SM: Adding Objects : blue	3.11	0.51	0.98	−0.49	−0.09	1.00
SM: Adding Objects : yellow	2.44	0.40	0.99	9.15	1.60	0.38
SM: Adding Objects : green	0.46	0.07	1.00	−4.73	−0.83	0.88
SM: Adding Objects : red	−6.00	−0.98	0.80	−3.93	−0.69	0.93
SM: Changing Refl. : blue	11.57	2.00	0.18	10.63	1.75	0.29
SM: Changing Refl. : yellow	3.83	0.66	0.94	3.73	0.61	0.96
SM: Changing Refl. : green	−12.36	−2.13	0.13	−6.26	−1.03	0.77
SM: Changing Refl. : red	−3.04	−0.53	0.97	−8.11	−1.33	0.56
Contrasts within cue (DL)	**Estimate**	**t-stat**	**p-val**	**Estimate**	**t-stat**	**p-val**
Local Surround : daylight	2.99	0.95	0.57	7.55	2.27	0.049
Maximum Flux : daylight	1.06	0.34	0.93	−3.22	−0.98	0.55
SM: Adding Objects : daylight	2.77	0.79	0.68	4.33	1.27	0.37
SM: Changing Refl. : daylight	7.70	2.32	0.04	7.18	1.99	0.10

Table 1. Analysis results of linear mixed-models fit to data (indoor and outdoor scenes separately). The first model was fitted with *cue* as a factor, *observer* as a random-effects variable, and a forced intercept of 0. The upper part displays the CCI difference estimate (departure from 0, i.e., baseline), *t-stat* for each silenced cue, and rounded *p*-values. The degrees of freedom was 135 for the indoor scene and 132 for the outdoor. The middle part shows the results of contrasts comparing illuminants within each silenced cue (after fitting the same LMM but with the interaction of *cue* and *illuminant* instead as factors). The lower part shows the results of contrasts comparing daylight and off-daylight illuminants within each cue. Estimates are estimated marginal mean differences from the mean CCI of the silenced cue. Degrees of freedom were 120 for all *t*-tests on contrasts. Tests with statistical significance ($p < 0.05$) are highlighted in green. See text for more details.

(Tukey, 1949), blue vs. green: $t(26) = -3.62$, $p = 0.007$; red vs. green: $t(26) = -2.9$, $p = 0.04$; yellow vs. green: $t(26) = -2.71$, $p = 0.05$). The high constancy under green illumination may be due to the overall green shift in this landscape scene, which might shift as well the perceived neutral.

Additionally, we wanted to compare illuminants along the daylight axis (blue and yellow) with those off the axis (red and green). We fit another linear mixed-effects model with *daylight* as a fixed-effects factor and *observer* as a random-effects factor and performed a one-way ANOVA. We did not find a

significant difference between illuminants on and off the daylight axis for the indoor scene ($F(1, 29) = 0.80$, $p = 0.38$), but we did for the outdoor—off-daylight illuminants resulted in higher CCIs than daylight illuminants (average CCIs of 92.7% and 77.3%, respectively; $F(1, 28) = 4.96$, $p = 0.03$).

Lastly, we combined both the indoor and outdoor data into another linear mixed-effects model with *scene* as a factor and *observer* again as a random-effects variable and performed a one-way ANOVA. We found no significant effect of scene ($F(1, 18) = 2.97$, $p = 0.10$), but there was a trend in the direction of better constancy in the indoor scene (96.7% vs. 85.2%).

For each scene, we fitted a linear mixed-effects model to the difference in color constancy indices between the baseline condition and each cue-silencing condition per observer using *cue* as a factor, *observer* as a random-effects variable, and a forced intercept of 0 (because we wanted to investigate departures from the baseline condition). We performed post hoc testing by fitting another linear model with *cue* and *illuminant* as factors and again with *observer* as a random-effects variable and a forced intercept of 0. We computed the estimated marginal means and applied contrasts with the method "eff," grouping by cue. Table 1 compiles the results of all tests on the models.

Effect of local surround

Figure 13 shows the effect of silencing the local surround cue for both the indoor and the outdoor environment. For the indoor scene, we found significantly reduced color constancy compared to baseline, with an average drop of 22.6%, from 96.7% to 74.1% ($t(135) = -4.89$, $p < 0.001$). In the outdoor scene, we also found significantly reduced constancy (average drop: 14.9%, from 84.7% to 69.8%; $t(132) = -4.14$, $p < 0.001$). A closer look at the contributions of each illuminant resulted in no significant difference between illuminants in the indoor scene, but in the outdoor scene, we did find that the green illuminant had a significantly greater effect than the other illuminants (average percentage drop for green: 38.1%; average percentage drop for red, blue, and yellow: 6.58%; $t(120) = -4.32$, $p < 0.001$, with Dunn–Šidák corrections; Šidák, 1967).

Effect of maximum flux (brightest region in image)

Figure 14 shows the effect of silencing the brightest region in the scene (maximum flux) for both the indoor and the outdoor environments. For both scenes, we did not find a significant difference in constancy indices (indoor: $t(135) = -1.47$, $p = 0.14$; outdoor: $t(132) = -0.24$, $p = 0.81$). The mean percentage drop in constancy was 6.8% (from 96.7% to 89.9%) and 0.9% (from 85.2% to 84.4%) for the indoor and outdoor scenes, respectively.

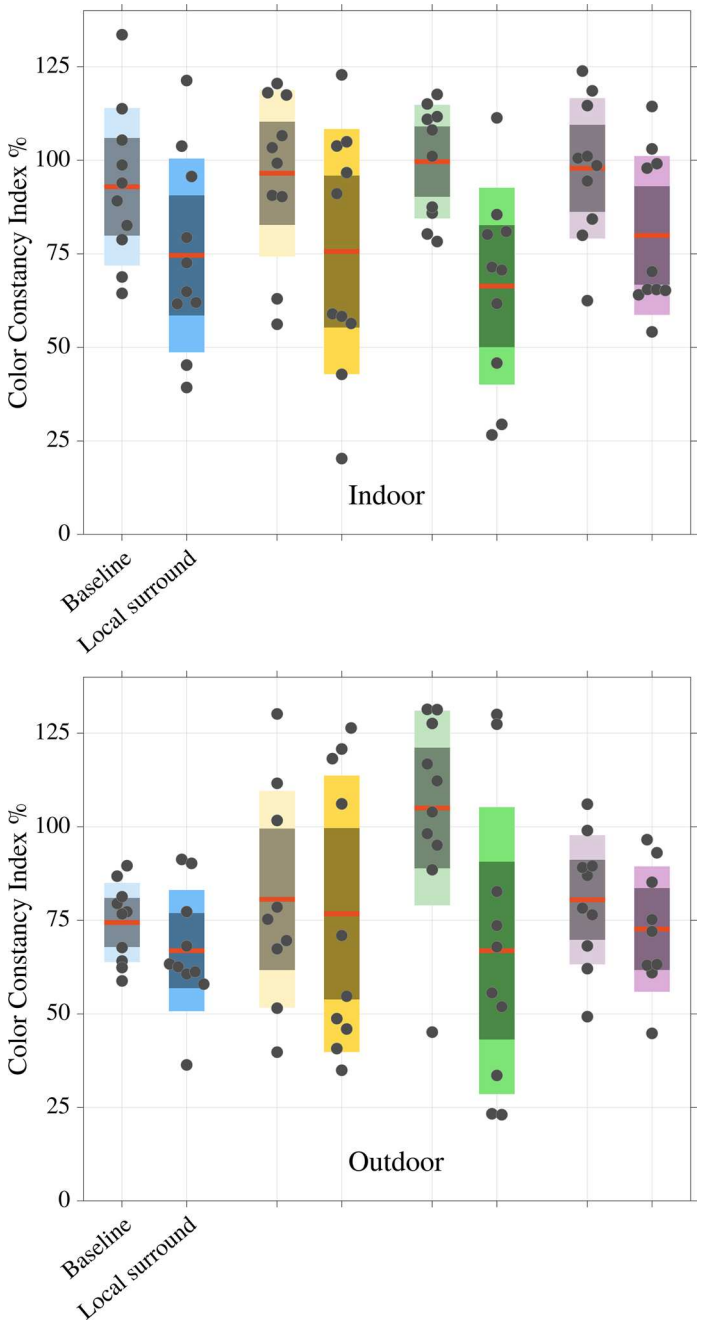

Figure 13. Color constancy indices when silencing the local surround cue along with indices for the baseline measure in the indoor (top) and outdoor (bottom) environments. The results are grouped by illuminant (color-coded), with the baseline results on the left and the cue-silencing results on the right. All other markings are the same as in Figure 12. Silencing the local surround cue had a significant effect on color constancy in both scenes ($p < 0.001$). Post hoc contrasts showed that the green illuminant in the outdoor scene reduced constancy more than the other illuminants ($p < 0.001$).

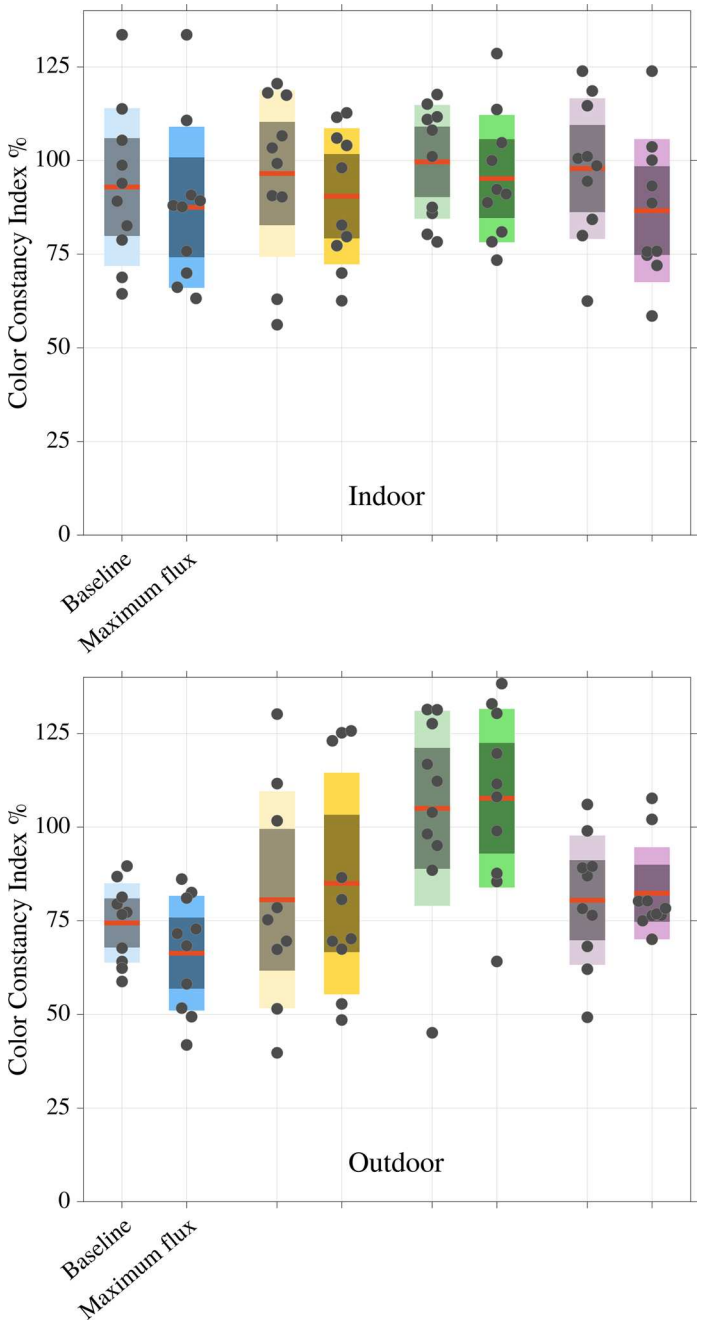

Figure 14. Color constancy indices after silencing the maximum flux cue in the indoor and outdoor scenes, along with baseline results. Markings are the same as in Figure 13. Color constancy after silencing this cue was not significantly different from the baseline measurements.

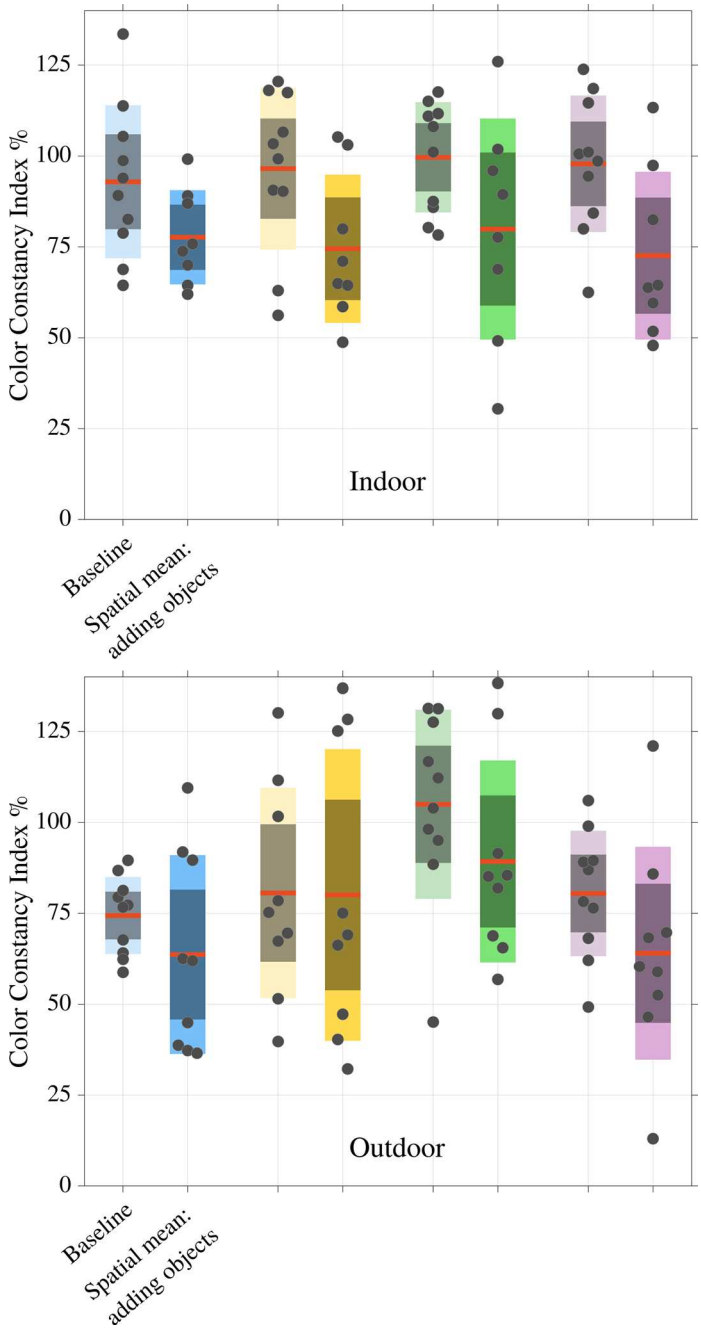

Figure 15. Color constancy indices after silencing the spatial mean color by adding objects, for both the indoor and outdoor environments, along with baseline results. Markings are the same as in Figures 13 and 14. Silencing the spatial mean cue in this way significantly reduced color constancy indices ($p \leq 0.01$).

Effect of spatial mean color (adding objects)

Figure 15 shows the effect of silencing the spatial mean color by adding new objects to the scene. In both scenes, we found a significant reduction in constancy comparable to the effect of silencing the local surround cue (indoor: $t(135) = -4.10$, $p < 0.001$; outdoor: $t(132) = -2.61$, $p = 0.01$). On average, the percentage reduction in color constancy index was 20.6% (from 96.8% to 76.2%) in the indoor scene and 9.7% (from 84.0% to 74.3%) in the outdoor scene.

Effect of spatial mean color (changing reflectances)

Figure 16 shows the effect of silencing the spatial mean color by changing the reflectances of objects

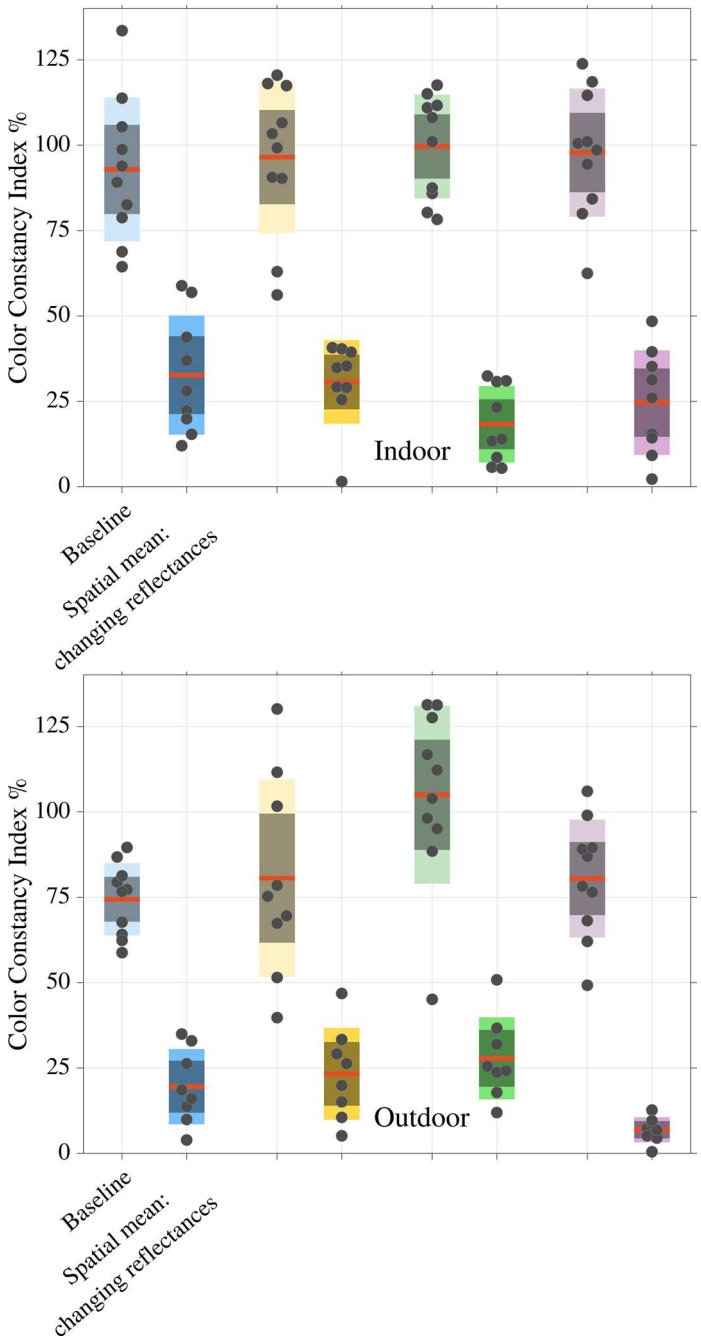

Figure 16. Color constancy indices after silencing the spatial mean color by changing object reflectances, along with baseline results. Markings are the same as in Figures 13 to 15. Silencing the cue in this way greatly affected constancy ($p < 0.001$).

already present in the scene. We see that silencing the cue in this way results in almost no color constancy: The average drop in constancy percentage was 67.3% (from 93.9% to 26.6%) and 66.1% (from 85.6% to 19.4%) in the indoor and outdoor scenes, respectively (indoor: $t(135) = -14.0$, $p < 0.001$; outdoor: $t(132) = -16.9$, $p < 0.001$).

Summary

Figure 17 gives an overall sense of the effects on color constancy of silencing each mechanism by plotting the percentages from each cue condition averaged across illuminant and scene. We can see that the outcome of each cue silencing is comparable across illuminants and scenes. Silencing the spatial mean cue by adding objects has a similar, mild effect on color constancy as silencing the local surround mechanism, while silencing the spatial mean cue by changing the reflectances of objects already present in the scene reduces constancy to approximately 25%. The maximum flux cue hardly affects constancy indices.

Differences between daylight and off-daylight illuminants

We explored comparisons between daylight and off-daylight illuminants within each silenced cue. For all cues except Maximum Flux in the outdoor scene, the mean drop in constancy was higher for off-daylight than daylight illuminants; however, this difference was significant only for the Local Surround cue in the outdoor scene and the Spatial Mean: Changing Reflectances cue in the indoor scene (see statistics in Table 1). Collapsing across all cues, we find that the off-daylight illuminants were more impacted by the cue silencing (indoor: $F(2, 137) = 24.9$, $p < 0.001$; outdoor: $F(2, 134) = 32.0$, $p < 0.001$). For the indoor scene, the average drop in constancy for daylight illuminants was 25.2% and for off-daylight illuminants was 32.3%. For the outdoor scene, the average drop was 17.5% for daylight illuminants and 24.7% for off-daylight.

Individual differences

We calculated the Pearson correlation coefficient of the constancy indices between observers for each scene. Observers were generally well correlated with each other, with an average correlation coefficient of 0.62 for the indoor scene and 0.51 for the outdoor scene. Observers 5 and 10 in the indoor scene and Observer 19 in the outdoor scene differed the most from the other observers. Figure A.4 in the Appendix explores differences among the neutral matches under all cue conditions; observers are largely consistent.

Discussion

Our results confirm previous findings that color constancy can be quite high in natural, immersive

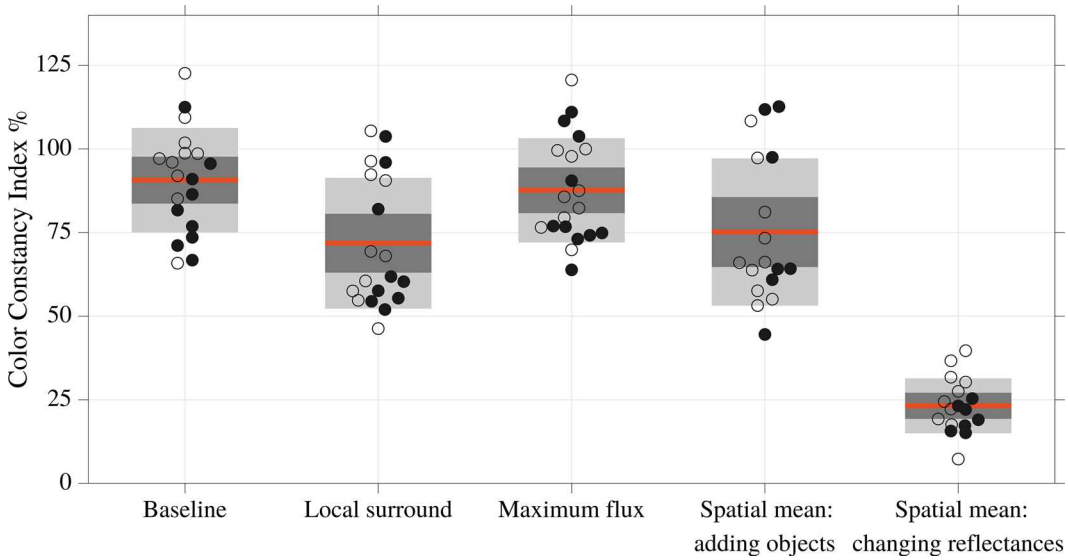

Figure 17. Color constancy indices per cue condition, averaged across illuminants and scenes. Open circles are participants from the indoor scene and filled circles from the outdoor scene. All other markings are as in Figures 13 to 16.

settings. Our investigation of three basic mechanisms for color constancy has resulted in a more elaborate picture on the importance of these mechanisms. Local context plays an important role, especially when context and illumination color interact with each other. Maximum flux, or "bright is white," did not have much of an effect in our settings, which might indicate that the relevance of the brightest object might play a role. The importance of global average color, or the "gray world assumption," depends on the scene. When all objects are corrected to keep mean color constant under an illumination change, color constancy breaks down. When new objects are embedded in an existing scene, they have a much reduced effect on constancy. All of these findings show that it is useful to study color constancy in naturalistic settings and that VR technology is a suitable tool to do so.

Color constancy mechanisms

Color constancy is a particularly interdisciplinary field. It has long been studied for basic vision research and became of great interest with the widespread use of color photography and film. With the advent of smartphones, probably more photographs and videos are taken than ever before, and proper color balancing is essential for viewing images under different lighting contexts. Thus, a multitude of algorithms for this task have been developed that allow estimating and compensating for illuminant effects, commonly known as "white balance" (Barnard et al., 2002; Gijsenij & Gevers, 2011; Gao, Yang, Li, & Li, 2015; Akbarinia & Parraga, 2018; Ulucan, Ulucan, & Ebner, 2022). Most recently, deep neural networks were introduced for color constancy (Lou, Gevers, Hu, & Lucassen, 2015; Akbarinia & Gil Rodríguez, 2020; Xu, Liu, Hou, Liu, & Qiu, 2020; Flachot et al., 2022; Heidari-Gorji & Gegenfurtner, 2023) and for illuminant estimation, even in challenging multi-illuminant scenarios (Li, Wang, Brown, & Tan, 2022). There is good agreement in basic and applied research that the three cues we investigated here are of utmost importance for man and machine.

We used VR to evaluate color constancy in an indoor and outdoor scene. This allowed us to achieve the high degrees of constancy observed in previous real-world immersive experiments (Kraft & Brainard, 1999; Mizokami, Ikeda, & Shinoda, 2004; Olkkonen et al., 2010; Pearce et al., 2014; Morimoto, Mizokami, Yaguchi, & Buck, 2017; Gupta, Gross, Pastilha, & Hurlbert, 2020; Gegenfurtner et al., 2024). Our results demonstrate the advantages of using virtual reality for such studies: the ability to have good cue control, to reproduce colors with high accuracy, and to manipulate scenes in real time, all while preserving the naturalism and immersion of the scenes. In the following sections, we will discuss our findings for the three mechanisms separately.

Local surround

The effect of the local surround cue is well supported by previous literature (von Kries, 1905; Land & McCann, 1971; Land, 1986; Valberg & Lange-Malecki, 1990; Foster & Nascimento, 1994), and early color constancy algorithms typically calculated the cone-excitation ratios between an object and surrounding areas to account for it. The most

important new aspect of our results is that there are interactions between illumination color and the color of the local surround. There was an overall effect of silencing the local surround cue (see Figure 13), but the effect was greater under the green illuminant. This was significant for the outdoor scene, and there was a strong trend in the indoor scene.

The particular effect of this cue on the green illuminant is likely because we used a rose-colored surround. This color is most opposite to the green illuminant in an opponent color space and thus leads to a color-contrast effect: The greenish lizard competitors appear even greener on the rose-colored leaves and so observers perceive one of the middle competitors to be more like the "Lonely Lizard" reference (Chevreul, 1861; Helmholtz, 1867; Kirschmann, 1892). These important interactions between local contrast and global illumination need to be explored in further experiments. It is also of interest that our modest manipulation of silencing the immediate surround already led to a noticeable drop in CCIs. It is unknown what extent of an object's surround might still have an impact on its perceived color, as well as the chromatic variation in that surround.

Maximum flux

Interestingly, the maximum flux cue (Land & McCann, 1971) did not have any significant effect on color constancy indices in our study, despite being frequently used in computational color constancy algorithms (Rizzi, Gatta, & Marini, 2002; Ebner, 2003; Garud, Ray, Mahadevappa, Chatterjee, & Mandal, 2014) and as a popular benchmark for those computational methods. While this cue has been studied for its role in lightness constancy (Cataliotti & Gilchrist, 1995; Gilchrist et al., 1999; Sharan, Li, Motoyoshi, Nishida, & Adelson, 2008; Anderson, Whitbread, & de Silva, 2014; Zeiner & Maertens, 2014), not as much psychophysical work exists on its role in color constancy. Kraft and Brainard (1999) observed a marked decrease in constancy to an average of 33% when the maximum flux cue was silenced. However, in their case, the region of maximum flux was also the region surrounding the text patch. Thus, the local surround cue was also affected by their manipulation. Linnell and Foster (2002) found maximum flux to be a better cue for color constancy than spatial mean only when Mondrian patches were no smaller than 1° (although in their experiments, observers made illumination matches, not object matches). Golz and MacLeod (2002) argued that not only the brightest region is important for judging illumination, but that more generally, a correlation between luminance and a particular hue (e.g., "bright is red") might be used by the visual system. In our experimental design, only the brightest object is affected by our silencing operation, but not any potential correlation of the illuminant with the luminance of a particular hue.

Our results do show a trend toward a contribution of maximum flux, and there might be several reasons why it did not play a more important role. In the indoor scene, we had chosen the lamp as the brightest object whose chromaticity we held constant across illuminations. The effect should be that the lamp would always appear as bright and white, anchoring the perception of all other colors in the scene. Instead, it seems that other cues such as the global average color determined perceived illumination, and the lamp's color was perceived as chromatically biased. The advent of light sources with largely variable spectra might support this interpretation of the scene. In the outdoor scene, we introduced a new object to the scene, a blanket sitting on the large cliff. While the blanket's size was larger than the lamp's, it still did not have any effect on the degree of constancy. To solve this riddle, we plan to run further experiments systematically exploring the size of the brightest object and ways to give that object a more prominent role in the scene.

It is also possible that the restricted dynamic range of the VR headset prevented a clear effect of maximum flux. The OLED screen does have an exquisite dynamic range, but the luminance difference between the brightest object and other bright objects might not have been large enough. In the natural world, the brightest regions are often specular reflections of the light source (Hurlbert, 1998), and these would be orders of magnitude brighter than other objects. However, they cannot readily be displayed on computer screens and are challenging even for high dynamic range systems. Newer headsets are now capable of providing a wider dynamic range and higher luminance values of up to 5,000 cd/m^2.

Spatial mean color

This classic cue, also known as the "gray world" hypothesis, postulates that humans account for the illuminant by recalibrating the average color of the scene to gray (Buchsbaum, 1980), thus eliminating the contribution of the illuminant. While there are known counterexamples to this account of color constancy (Gilchrist & Jacobsen, 1984; Bäuml, 1994; Jenness & Shevell, 1995; Brown & MacLeod, 1997; Webster & Mollon, 1997; Bloj, Kersten, & Hurlbert, 1999; Kraft & Brainard, 1999; Golz & MacLeod, 2002; Ennis & Doerschner, 2019), it is probably still considered the most important cue. Our results supplement earlier research by suggesting that the spatial mean color computation is dependent on some sort of segmentation of the scene, rather than a pixel-wise computation. The effect of grouping and segmentation on perception is not new (Anderson & Winawer, 2005;

Werner, 2006; Gilchrist & Radonjić, 2010; Ennis & Doerschner, 2019); however, it is not clear exactly what kind of segmentation observers might be using. A simple partitioning of foreground versus background is likely not the explanation. One might say that observers could have segmented the outdoor scene into the foreground's floating, "moon-like" objects and the background's more natural objects and then discard the foreground objects in order to be color constant, but this would be more difficult for the indoor scene, where the added objects are affixed and not easily separable from the rest of the scene. Other work suggests that scenes might be separable based on spatial frequency bands (high- vs. low-frequency content, the latter of which the spatial mean would fall under) (Dixon & Shapiro, 2017). However, at least along the luminance and chroma dimensions (see Scene statistics in Appendix), we find little differences in the spatial frequency content between the two scenes. Significant differences possibly exist along more relevant dimensions, such as hue, but a thorough quantification of these statistics and systematic exploration of their relationship is beyond the scope of this study.

One significant aspect of the Changing Reflectances cue silencing is that almost all of the static objects changed between illuminant scenes. This means we are essentially testing if observers are color constant in an ever-changing environment. One could argue that this would result in poor color constancy because observers recognized that the structure in each scene is the same and thus expected that the reflectances of the objects are exactly the same across illuminants. But we know that humans are largely color constant when a given object is moved around to new scenes with different lighting. One way to test this would be to change noticeable parts of the scene under each illuminant such that the observers would recognize the scenes as different and eliminate any assumption that the scenes, and therefore object reflectances, are the same. Alternatively, observers may instead have had some prior idea about possible object reflectances in an office scene or a forest scene, whether due to some calculation of the optimal colors under a given illuminant (Morimoto, Kusuyama, Fukuda, & Uchikawa, 2021) or memory colors of certain recognizable objects (Granzier & Gegenfurtner, 2012) or of schemas or gists of scenes (Castelhano & Henderson, 2008). Thus, the reflectances chosen to balance the spatial mean color under some of the illuminants may have been perceived as unlikely to occur. One could test this by creating abstract scenes and similarly manipulating reflectances under each illuminant. But results from Gilchrist and Ramachandran (1992; see also Ruppertsberg & Bloj, 2007; Langer, 1999) suggest at least that color constancy would not be entirely eliminated. They asked, "Is a red room under white light discriminable from a white room under red light?" and found that observers could distinguish the rooms due to the interreflections in the shadows.

It is important to note that, in our two implementations of silencing the spatial mean cue, the local surround cue was affected differently. In the Adding Objects scenes, the lizards were never placed on the added objects. Thus, with the exception of added interreflections, there was little difference between the immediate surround of the lizards in these scenes and in the baseline scenes. Alternatively, in the Changing Reflectances scenes, almost all of the objects on which the lizards sat were different compared to the Baseline scene, and as a whole, these objects encompassed almost the entire scene. This means that long-range as well as local contrast between lizard and surround was different between the two spatial mean scenes and may have affected constancy. Supporting this, we can see from Figure A.2 in the Appendix that the Adding Objects scenes contained a wider range of colors than the Changing Reflectances scenes. Several studies have shown that the chromatic variability surrounding an object can have varied effects on the object's color appearance, even if the average color of the surround is equal (Jenness & Shevell, 1995; Brown & MacLeod, 1997; Mausfeld & Andres, 2002). Results from Golz and MacLeod (2002) suggest that the correlation between luminance and chromaticity across a scene (specifically redness) can be diagnostic of the illuminant (see, however, Ciurea & Funt, 2004; Granzier, Brenner, Cornelissen, & Smeets, 2005). Additionally, the location of the added objects in relation to the lizards might greatly affect local adaptation on the retina—since the added objects, which the lizards were never placed on top of, were otherwise irrelevant. Observers looked at the frames and ivy plants in the indoor scene, with frequencies ranging from 12% to 24%. We plan to further examine individual variability by comparing observer responses with their eye movements around the scenes.

Conclusions

Our results show that VR is a powerful method to study color constancy in natural, photo-realistic scenes. Not only can high levels of constancy be achieved, but the scenes can be much more easily manipulated and cues to constancy can be experimentally investigated in ways not possible in real-world setups. This paves the way for experiments with stimuli having more complex material properties and experiments with multiple illuminations within one scene. At the conceptual level, our results are based on the important experiments by Kraft and Brainard (1999), which we replicate and extend. Our results show that color constancy is not pixel based but rather depends on the objects within a scene and their interaction with the illumination color.

Keywords: color constancy, local surround, spatial mean, maximum flux, virtual reality

Acknowledgments

Supported by ERC Advanced Grant Color 3.0 (Grant no. 884116 and BBSRC Grant BB/X01312X/1). This research was also supported by "The Adaptive Mind," funded by the Excellence Program of the Hessian Ministry of Higher Education, Research, Science and the Arts.

Commercial relationships: none.
Corresponding author: Raquel Gil Rodriguez.
Email: raquel.gil-rodriguez@psychol.uni-giessen.de.
Address: Psychology Department, Justus-Liebig University Giessen, Alter Steinbacher Weg 38, 35394 Giessen, Germany.

*RGR and LH are contributed equally to this work.

References

Akbarinia, A., & Gil Rodríguez, R. (2020). Deciphering image contrast in object classification deep networks. *Vision Research, 173*, 61–76.

Akbarinia, A., & Parraga, C. A. (2018). Colour constancy beyond the classical receptive field. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(9), 2081–2094.

Anderson, B. L., Whitbread, M., & de Silva, C. (2014). Lightness, brightness, and anchoring. *Journal of Vision, 14*(9), 7, https://doi.org/10.1167/14.9.7.

Anderson, B. L., & &Winawer, J. (2005). Image segmentation and lightness perception. *Nature, 434*(7029), 79–83.

Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A, 3*(10), A262–A270.

Aston, S., Radonjić, A., Brainard, D. H., & Hurlbert, A. C. (2019). Illumination discrimination for chromatically biased illuminations: Implications for color constancy. *Journal of Vision, 19*(3), 15, https://doi.org/10.1167/19.3.15.

Barnard, K., Cardei, V., & Funt, B. (2002). A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing, 11*(9), 972–984.

Bloj, M. G., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature, 402*, 877–879.

Brainard, D. H., & Wandell, B. A. (1986). Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A, 3*(10), 1651–1661.

Brown, R. O., & MacLeod, D. I. A. (1997). Color appearance depends on the variance of surround colors. *Current Biology, 7*(11), 844–849.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute, 310*(1), 1–26.

Bäuml, K.-H. (1994). Color appearance effects of illuminant changes under different surface collections. *Journal of the Optical Society of America A, 11*(2), 531–542.

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance, 34*(3), 660–675.

Cataliotti, J., & Gilchrist, A. L. (1995). Local and global processes in surface lightness perception. *Perception & Psychophysics, 57*, 125–135.

Chevreul, M. E. (1861). *The laws of contrast of colour: And their application to the arts*. London: Routledge, Warne, and Routledge.

Cipresso, P., Wilson, C. J., & Soranzo, A. (2015). The use of virtual reality in psychology: A case study in visual perception. *Computational and Mathematical Methods in Medicine, 2015*, 7.

Ciurea, F., & Funt, B. (2004). Failure of luminance-redness correlation for illuminant estimation. In *Proceedings of the 12th Color and Imaging Conference* (pp. 42–46). Springfield, VA: Society for Imaging Science and Technology.

Clausen, O., Fischer, G., Furhmann, A., & Marroquim, R. (2019). Towards predictive virtual prototyping: Color calibration of consumer VR HMDs. In *16th GI AR/VR Workshop, Fulda, Germany* (pp. 13–24).

Dixon, E. L., & Shapiro, A. G. (2017). Spatial filtering, color constancy, and the color-changing dress. *Journal of Vision, 17*(3), 7, https://doi.org/10.1167/17.3.7.

Díaz-Barrancas, F., Cwierz, H., Gil Rodríguez, R., & Pardo, P. J. (2022). Validating perception of hyperspectral textures in virtual reality systems. In M. Krone, S. Lenti, & J. Schmidt (Eds.), *EuroVis 2022 – Posters* (pp. 51–53). Rome, Italy: The Eurographics Association.

Díaz-Barrancas, F., Gil Rodríguez, R., Aizenman, A., Bayer, F., & Gegenfurtner, K. R. (2023). Color calibration in virtual reality for unity and unreal. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Shanghai, China* (pp. 733–734).

Ebner, M. (2003). Combining white-patch retinex and the gray world assumption to achieve color constancy for multiple illuminants. In B. Michaelis & G. Krell (Eds.), *Pattern recognition* (pp. 60–67). Berlin, Germany: Springer.

Ennis, R., & Doerschner, K. (2019). Disentangling simultaneous changes of surface and illumination. *Vision Research, 159*, 173–188.

Flachot, A., Akbarinia, A., Schütt, H. H., Fleming, R. W., Wichmann, F. A., & Gegenfurtner, K. R. (2022). Deep neural models for color classification and color constancy. *Journal of Vision, 22*(4), 17, https://doi.org/10.1167/jov.22.4.17.

Foster, D. H. (2011). Color constancy. *Vision Research, 51*(7), 674–700. (Vision Research 50th Anniversary Issue: Part 1)

Foster, D. H., & Nascimento, S. M. C. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings: Biological Sciences, 257*(1349), 115–121.

Gao, S.-B., Yang, K.-F., Li, C.-Y., & Li, Y.-J. (2015). Color constancy using double-opponency. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(10), 1973–1985.

Garud, H., Ray, A. K., Mahadevappa, M., Chatterjee, J., & Mandal, S. (2014). A fast auto white balance scheme for digital pathology. In *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Valencia, Spain* (pp. 153–156), https://doi.org/10.1109/BHI.2014.6864327.

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology, 10*(13), 805–808.

Gegenfurtner, K. R., Weiss, D., & Bloj, M. (2024). Color constancy in real-world settings. *Journal of Vision, 24*(2), 12, https://doi.org/10.1167/jov.24.2.12.

Gijsenij, A., & Gevers, T. (2011). Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(4), 687–698.

Gilchrist, A. L., & Jacobsen, A. (1984). Perception of lightness and illumination in a world of one reflectance. *Perception, 13*(1), 5–19.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., … Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review, 106*(4), 795–834.

Gilchrist, A. L., & Radonjíc, A. (2010). Functional frameworks of illumination revealed by probe disk technique. *Journal of Vision, 10*(5), 6, https://doi.org/10.1167/10.5.6.

Gilchrist, A. L., & Ramachandran, U. (1992). Red rooms in white light appear different from white rooms in red light. *Investigative Ophthalmology and Visual Science, 33*, 756.

Gil Rodríguez, R., Bayer, F., Toscani, M., Guarnera, D., Guarnera, G. C., & Gegenfurtner, K. R. (2022). Colour calibration of a head mounted display for colour vision research using virtual reality. *SN Computer Science, 3*, 1–10.

Golz, J., & MacLeod, D. I. A. (2002). Influence of scene statistics on colour constancy. *Nature, 415*(6872), 637–640.

Granzier, J. J., Brenner, E., Cornelissen, F. W., & Smeets, J. B. (2005). Luminance—Color correlation is not used to estimate the color of the illumination. *Journal of Vision*, *5*(1), 2, https://doi.org/10.1167/5.1.2..

Granzier, J. J. M., & Gegenfurtner, K. R. (2012). Effects of memory colour on colour constancy for unknown coloured objects. *i-Perception, 3*(3), 190–215.

Gupta, G., Gross, N., Pastilha, R., & Hurlbert, A. (2020). The time course of colour constancy by achromatic adjust ment in immersive illumination: What looks white under coloured lights? *bioRxiv,* 2020.03.10.984567.

Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision, 7*(4), 2, https://doi.org/10.1167/7.4.2.

Heidari-Gorji, H., & Gegenfurtner, K. R. (2023). Object-based color constancy in a deep neural network. *Journal of the Optical Society of America A, 40*(3), A48–A56.

Helmholtz, H. v. (1867). *Handbuch der physiologischen optik*. Leipzig, Germany: Allgemeine Encyklopädie der Physik.

Helmholtz, H. v. (1910). A. Gullstrand, J. von Kries & W. Nagel (Eds.), *Handbuch der physiologischen optik* (3rd ed., *Vol. 3*). Hamburg: Vos.

Hering, E. (1920). *Grundzüge der lehre vom lichtsinn*. Berlin, Germany: Springer.

Hurlbert, A. C. (1998). Computational models of color constancy. In V. Walsh & J. Kulikowski (Eds.), *Perceptual constancy: Why things look as they do* (pp. 283–322). Cambridge, UK: Cambridge University Press.

Hurlbert, A. C., Gupta, G., Gross, N., & Pastilha, R. (2019). Colour constancy measured by achromatic adjustment in immersive illumination. *Journal of Vision, 19*(10), 674–700, https://doi.org/10.1167/19.10.296.

Imaoka, Y., Flury, A., & Bruin, E. de. (2020). Assessing saccadic eye movements with head-mounted display virtual reality technology. *Frontiers in Psychiatry, 11*, 572938.

Jenness, J. W., & Shevell, S. K. (1995). Color appearance with sparse chromatic context. *Vision Research, 35*(6), 797–805.

Jensen, H. W. (2001). *Realistic image synthesis using photon mapping*. Natick, MA: A.k. Peters Ltd/CRC Press.

Judd, D. B. (1940). Hue saturation and lightness of surface colors with chromatic illumination. *Journal of the Optical Society of America, 30*(1), 2–32.

Katz, D. (1911). *Die erscheinungsweisen der farben und ihre beeinflussung durch die individuelle erfahrung*. Leipzig, Germany: Ambrosius Barth.

Kirschmann, A. (1892). Some effects of contrast. *The American Journal of Psychology, 4*(4), 542–557.

Knoblauch, K., & Maloney, L. (Eds.). (2012). *Modeling Psychophysical Data in R.* New York, NY: Springer

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences, 96*(1), 307–312.

Kries, J. von. (1905). *Handbuch der physiologie des menschen*. (W. Nagel Ed., Vol. *3*, pp. 109–282). Braunschweig, Germany: Vieweg und Sohn.

Kries, J. von. (1923). *Allgemeine sinnesphysiologie*. Leipzig, Vogel.

Land, E. H. (1986). Recent advances in retinex theory. *Vision Research, 26*(1), 7–21.

Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America, 61*(1), 1–11.

Langer, M. (1999). When shadows become interreflections. *International Journal of Computer Vision, 34*(2/3), 193–204.

Lenth, R. V. (2023). emmeans: Estimated marginal means, aka least-squares means [Computer software manual], https://CRAN.R-project.org/package=emmeans.

Li, S., Wang, J., Brown, M. S., & Tan, R. T. (2022). Transcc: Transformer-based multiple illuminant color constancy using multitask learning. arXiv preprint arXiv:2211.08772.

Linnell, K. J., & Foster, D. H. (2002). Scene articulation: Dependence of illuminant estimates on number of surfaces. *Perception, 31*(2), 151–159.

Lou, Z., Gevers, T., Hu, N., & Lucassen, M. P. (2015). Color constancy by deep learning. In *British Machine Vision Conference* (pp. 76–81).

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision, 3*(8), 5, https://doi.org/10.1167/3.8.5.

Mausfeld, R., & Andres, J. (2002). Second-order statistics of colour codes modulate transformations that effectuate varying degrees of scene invariance and illumination invariance. *Perception, 31*(2), 209–224.

Mizokami, Y., Ikeda, M., & Shinoda, H. (2004). Color constancy in a photograph perceived as a three-dimensional space. *Optical Review, 11*(4), 288–296.

Morimoto, T., Kusuyama, T., Fukuda, K., & Uchikawa, K. (2021). Human color constancy based on the geometry of color distributions. *Journal of Vision, 21*(3), 7, https://doi.org/10.1167/jov.21.3.7.

Morimoto, T., Mizokami, Y., Yaguchi, H., & Buck, S. L. (2017). Color constancy in two-dimensional and three-dimensional scenes: Effects of viewing methods and sur face texture. *i-Perception, 8*(6), https://doi.org/10.1177/2041669517743522.

Murray, R. F., Patel, K. Y., & Wiedenmann, E. S. (2022). Luminance calibration of virtual reality displays in unity. *Journal of Vision, 22*(13), 1, https://doi.org/10.1167/jov.22.13.1.

Nascimento, S. M. C., & Foster, D. H. (2000). Relational color constancy in achromatic and isoluminant images. *Journal of the Optical Society of America A, 17*(2), 225–231.

Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision, 10*(9), 16, https://doi.org/10.1167/10.9.16.

Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. (2014). Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. *PLoS ONE, 9*(2), 1–10.

Pinheiro, J., & Bates, D., & R Core Team. (2023). nlme: Linear and nonlinear mixed effects models [Computer software manual], https://CRAN.R-project.org/package=nlme.

R Core Team. (2023). R: A language and environment for statistical computing [Computer software manual], https://www.R-project.org/.

Radiant Vision Systems. (2023). ProMetric Software, https://www.radiantvisionsystems.com/products/prometric-software.

Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015). Color constancy in a naturalistic, goal-directed task. *Journal of Vision, 15*(13), 1–21, https://doi.org/10.1167/15.13.3.

Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015). Color constancy supports cross-illumination color selection. *Journal of Vision, 15*(6), 13, https://doi.org/10.1167/15.6.13.

Radonjić, A., Ding, X., Krieger, A., Aston, S., Hurlbert, A. C., & Brainard, D. H. (2018). Illumination discrimination in the absence of a fixed surface-reflectance layout. *Journal of Vision, 18*(5), 11, https://doi.org/10.1167/18.5.11.

Rizzi, A., Gatta, C., & Marini, D. (2002). Color correction between gray world and white patch. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human Vision and Electronic Imaging VII* (Vol. *4662*, pp. 367–375). San Jose, California, USA: SPIE.

Ruppertsberg, A. I., & Bloj, M. (2007). Reflecting on a room of one reflectance. *Journal of Vision, 7*(13), 12, https://doi.org/10.1167/7.13.12.

Schregle, R. (2003). Bias compensation for photon maps. *Computer Graphics Forum, 22*(4), 729–742.

Shapiro, A. G., & LoPrete, A. (2020). Helix rotation: Luminance contrast controls the shift from two-dimensional to threedimensional perception. *Journal of the Optical Society of America A, 37*(4), A262–A270.

Sharan, L., Li, Y., Motoyoshi, I., Nishida, S., & Adelson, E. H. (2008). Image statistics for surface reflectance perception. *Journal of the Optical Society of America A, 25*(4), 846–865.

Sharma, G., Wu, W., & Dalal, E. N. (2005). The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application, 30*(1), 21–30.

Ŝidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62*(318), 626–633.

Toscani, M., Gil Rodríguez, R., Guarnera, D., Guarnera, G. C., Kalouaz, A., & Gegenfurtner, K. R. (2019). Assessment of OLED head mounted display for vision research with virtual reality. In *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems, Sorrento, Italy* (pp. 738–745).

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics, 5*(2), 99–114, http://www.jstor.org/stable/3001913.

Ulucan, O., Ulucan, D., & Ebner, M. (2022). Color constancy beyond standard illuminants. In *2022 IEEE International Conference on Image Processing, Bordeaux, France* (pp. 2826–2830).

Valberg, A., & Lange-Malecki, B. (1990). 'Colour constancy' in Mondrian patterns: A partial cancellation of physical chromaticity shifts by simultaneous contrast. *Vision Research, 30*(3), 371–380.

Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology, 38*(3), 310–324.

Webster, M. A., & Mollon, J. D. (1997). Adaptation and the color statistics of natural images. *Vision Research, 37*(23), 3283–3298.

Werner, A. (2006). The influence of depth segmentation on colour constancy. *Perception, 35*(9), 1171–1184.

Wichmann, F., Sharpe, L., & Gegenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 509–520.

Wiesing, M., Fink, G. R., & Weidner, R. (2020). Accuracy and precision of stimulus timing and reaction times with Unreal Engine and SteamVR. *PLoS ONE, 15*, 4.

Witzel, C., & Gegenfurtner, K. R. (2018). Color perception: Objects, constancy, and categories. *Annual Review of Vision Science, 4*(1), 475–499.

Xu, B., Liu, J., Hou, X., Liu, B., & Qiu, G. (2020). End-to-end illuminant estimation based on deep metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3613–3622). Seattle, WA, USA.

Yaremych, H. E., & Persky, S. (2019). Tracing physical behavior in virtual reality: A narrative review of applications to social psychology. *Journal of Experimental Social Psychology, 85*, 103845.

Zaman, N., Sarker, P., & Tavakkoli, A. (2023). Calibration of head mounted displays for vision research with virtual reality. *Journal of Vision, 23*(6), 7, https://doi.org/10.1167/jov.23.6.7.

Zeiner, K., & Maertens, M. (2014). Linking luminance and lightness by global contrast normalization. *Journal of Vision, 14*(7), 3, https://doi.org/10.1167/14.7.3.

Appendix

Scene statistics

We compared the color distributions of the Baseline and two Spatial Mean scenes under each illuminant (with no lizards). We used the shots taken of the scene using the I29 imaging colorimeter (see Spatial mean: adding objects or changing reflectances) and, in addition to the mean CIExy chromaticities already calculated (Figure A.1), computed the mean in CIELAB

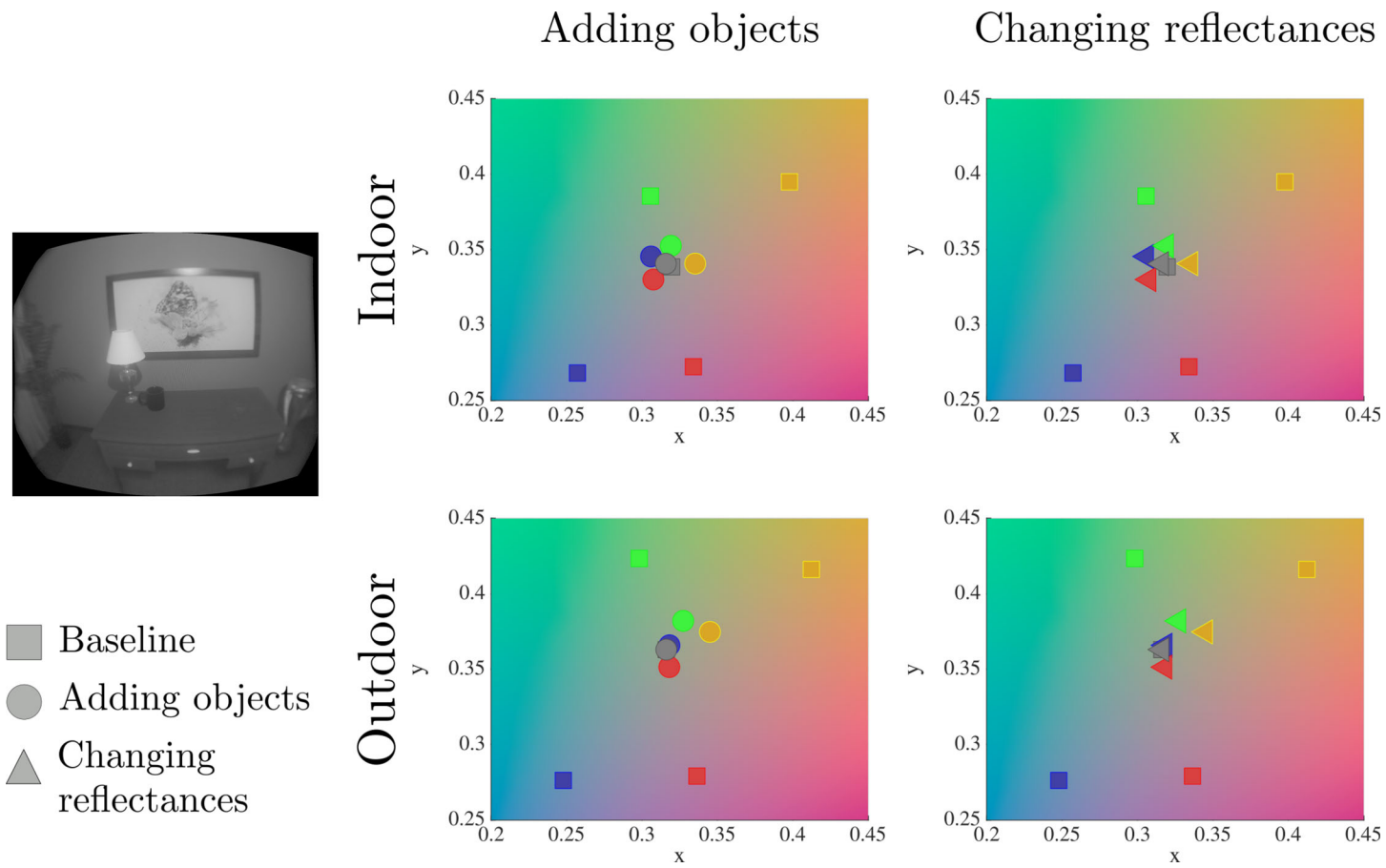

Figure A.1. Spatial average color across scenes for the Baseline (squares), Spatial Mean: Adding Objects (circles), and Spatial Mean: Changing Reflectances (triangles) conditions in CIExy chromaticities (top: indoor; bottom: outdoor). On the left is a shot taken by the colorimeter using the AR/VR lens of the indoor baseline scene under the yellow illuminant (luminance only shown). The average color for the baseline condition is plotted in all plots, respective of scene (squares color-coded by illuminant). The left plots show the spatial average color for the Spatial Mean: Adding Objects condition and the rightmost plots for the Spatial Mean: Changing Reflectances condition (circles and triangles, respectively, color-coded by illuminant). See text for details on measurement collection.

and plotted the distribution of pixels. Figure A.2 plots a representation of these color distributions (scaled for visibility—see figure caption). We can see that adding objects (squares) to shift the spatial mean color to neutral results in an elongation of the color distribution in the opponent color direction, but this elongation is not present in the Spatial Mean: Changing Reflectances scenes (triangles). We also see for many of the scenes that the distribution of colors is slightly larger for the outdoor scenes than the indoor scenes.

We also explored the spatial frequency content of various scenes in the luminance dimension and the chroma (CIELAB C*) dimension. Overall, we found negligible differences between scenes. We first compared the indoor and outdoor scenes under the neutral illuminant. While mean luminance and chroma were slightly different, we otherwise found very minute differences in all other spatial frequencies: Indoor mean luminance was 3.90 cd/m^2 greater than outdoor mean luminance, and the average magnitude difference for all other spatial frequencies was less than 0.001 cd/m^2 (standard deviation <0.001). Along the chroma dimension, the indoor scene had a mean chroma greater by 4.27; the average magnitude for all other spatial frequencies was less than 0.001 (standard deviation <0.001). We also compared the Adding Objects and Changing Reflectances scenes under the blue illuminant and again found negligible differences. In the indoor environment, we found a mean luminance difference of 0.96 cd/m^2 (Adding Objects was greater) and an average magnitude difference less than 0.001 cd/m^2 (standard deviation <0.001) for all other spatial frequencies. In the chroma dimension, Adding Objects had a mean value 5.13 points higher than Changing Reflectances; the average magnitude difference for all other spatial frequencies was also very low (mean < 0.001 C*; standard deviation = 0.002). In the outdoor environment, there was a mean luminance difference of 4.92 cd/m^2 (Adding Objects was greater) and an average magnitude difference less than 0.001 (standard

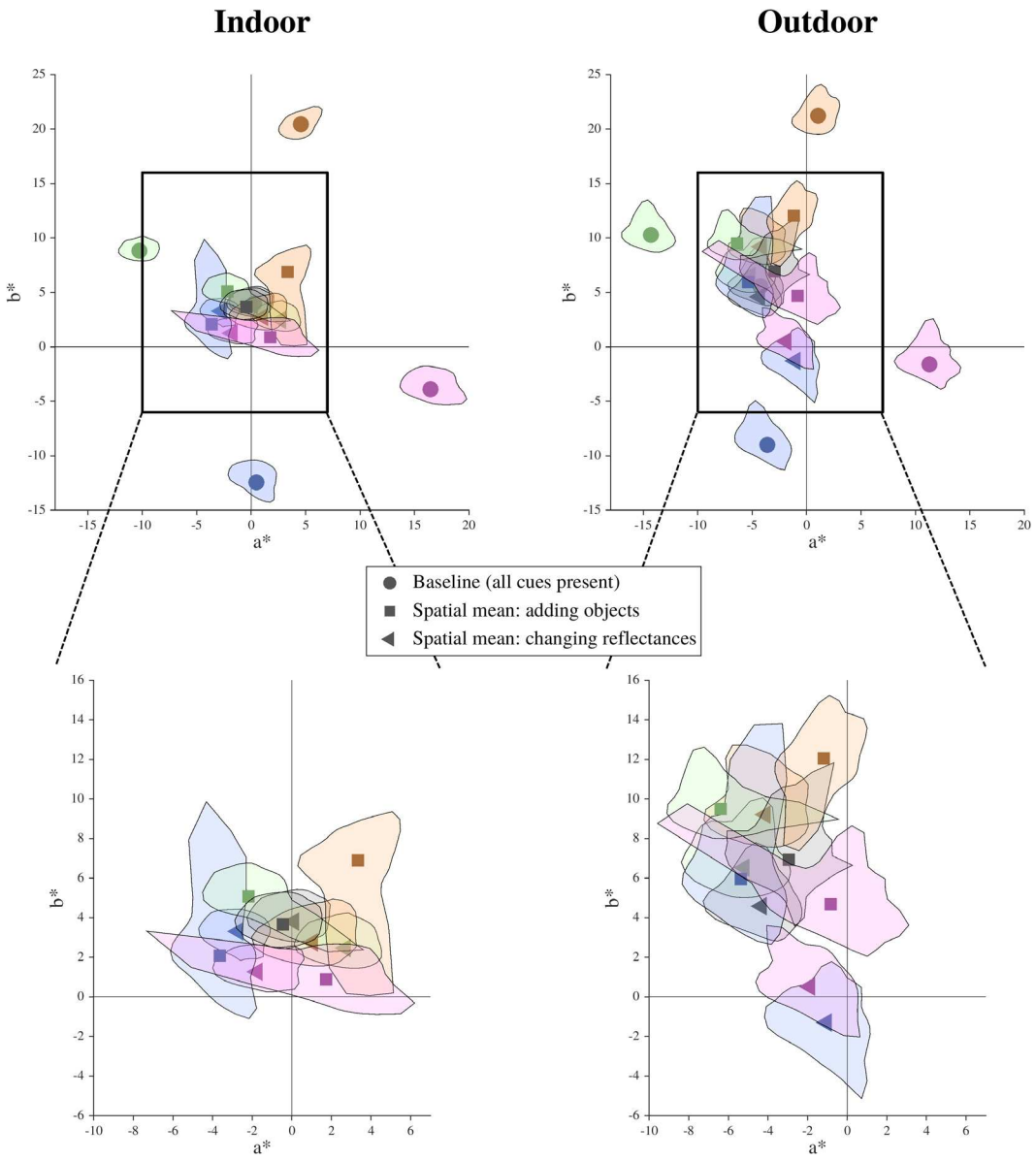

Figure A.2. Representation of the distribution of colors for the Baseline, Spatial Mean: Adding Objects, and Spatial Mean: Changing Reflectances cue conditions under each illuminant for the indoor and outdoor scenes, plotted in CIELAB. Markers represent the mean pixel color across each scene and are color-coded by illuminant. Marker shapes are as indicated in the legend. Shape of distribution was calculated by aggregating pixels from each of eight images taken by the colorimeter (see Spatial mean: adding objects or changing reflectances) and plotting them in CIELAB. Distance of each pixel color from the scene mean was calculated and averaged within 10° bins around scene mean. Plotted are these averages, scaled down by 5 to reduce clutter. Bottom plots are zoomed-in versions of top plots.

deviation = 0.002). For chroma, Adding Objects had a mean chroma higher by 3.59 points; for all other spatial frequencies, the average magnitude was again less than 0.001 (standard deviation = 0.002).

Results: Neutral illuminant

Measurements made under the neutral illuminant for each cue condition served as a control that the results obtained under the chromatic illuminants were not due to artifacts of the manipulations. We found no significant difference between the observers' matches under the neutral illuminant across all five conditions (one-way unbalanced ANOVA on a linear mixed-effects model with *cue* as a factor and *observer* as a random-effects variable; indoor: $F(4, 33) = 1.50$, $p = 0.22$; outdoor: $F(4, 33) = 0.47$, $p = 0.76$). Figure A.3 plots the matches in competitor space for each observer for the two scenes. One can see that observers were

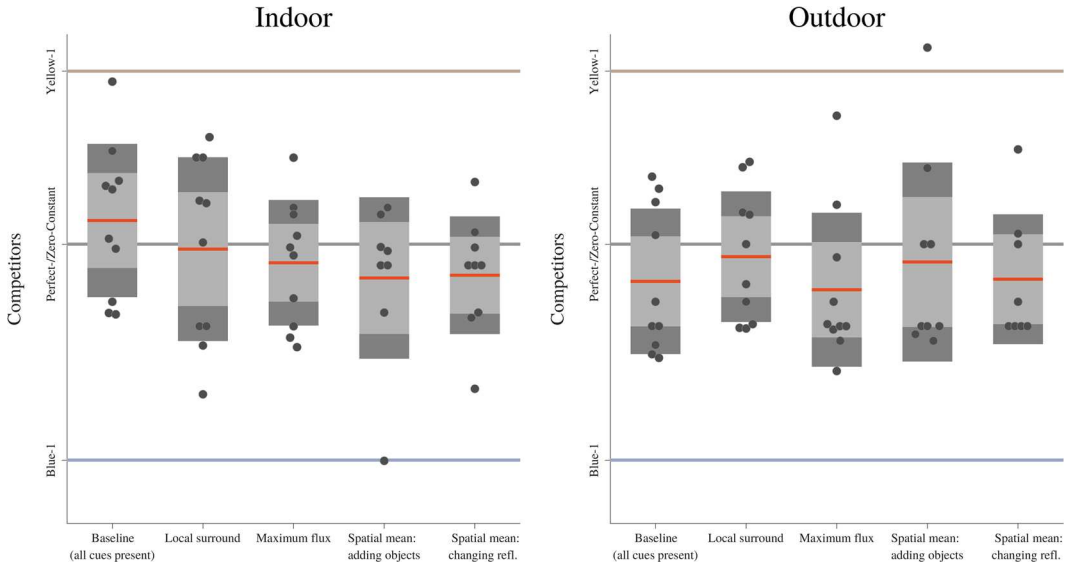

Figure A.3. Observer matches under the neutral illuminant for all cue conditions for both scenes. The location of each match with respect to the five competitors is plotted on the y-axis and jittered slightly along the x-axis to improve visibility. The darker boxed region denotes ±1.96 standard errors of the mean and the lighter region denotes ±1 standard deviation from the mean. The perfect-constant and zero-constant match are the same under the neutral illuminant and are marked here as a gray horizontal line. Two other competitors (one in the yellowish direction and one in the bluish direction) are also plotted as horizontal lines. Observer matches were calculated with an adapted MLDS algorithm (see Analysis section).

quite good at remembering and choosing the right lizard (the perfect-/zero-constant lizard). Note that when calculating the color constancy indices for the cue conditions, we recalculated the location of the perfect-constant and zero-constant competitors based on each observer's average match across the neutral conditions (see Analysis for details).

Average interobserver Pearson correlation coefficient was 0.03 for the indoor scene and 0.06 for the outdoor.

Figure A.4 shows the neutral match for each participant (x-axis) under the five conditions (color-coded: black, Baseline; blue, Local Surround; orange, Maximum Flux; purple, Spatial Mean: Adding Objects; green, Spatial Mean: Changing Reflectances) for both indoor and outdoor scenes. The y-axis plots three competitors, with the middle one representing the perfect-constant match.

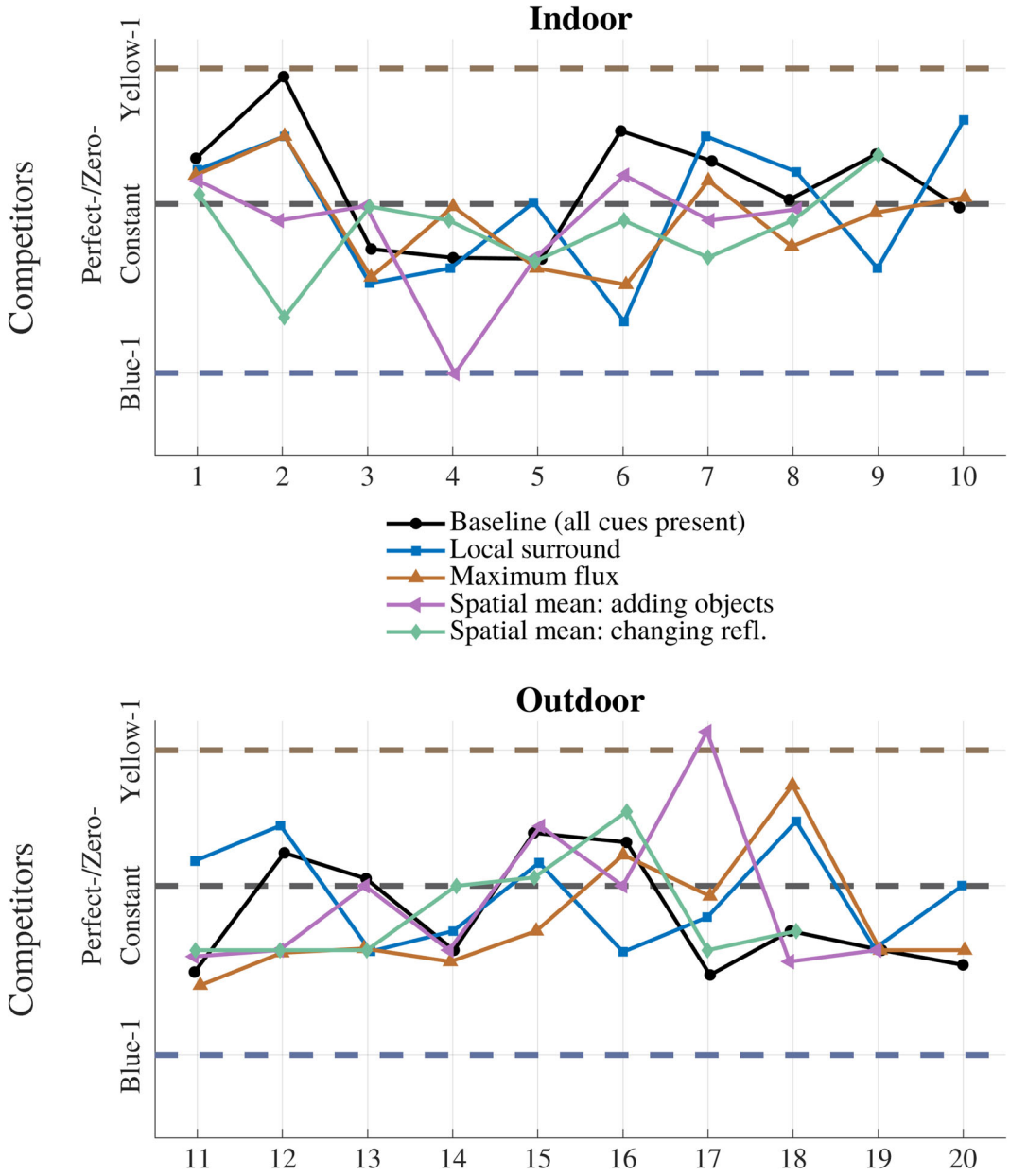

Figure A.4. Individual differences in matches under neutral illuminant for indoor (top) and outdoor (bottom) scenes. Observer IDs are along the x-axis (note that different sets of observers participated in the indoor vs. outdoor experiments). Along the y-axis are the locations of the match in competitor space, with the perfect-/zero-constant match and two competitors labeled. Matches for the different cue conditions are described in the legend and are connected with lines.